USING THE CENSORED GAMMA DISTRIBUTION FOR MODELING FRACTIONAL RESPONSE VARIABLES WITH AN APPLICATION TO LOSS GIVEN DEFAULT

BY

FABIO SIGRIST AND WERNER A. STAHEL

ABSTRACT

Regression models for limited continuous dependent variables having a non-negligible probability of attaining exactly their limits are presented. The models differ in the number of parameters and in their flexibility. Fractional data being a special case of limited dependent data, the models also apply to variables that are a fraction or a proportion. It is shown how to fit these models and they are applied to a Loss Given Default dataset from insurance to which they provide a good fit.

KEYWORDS

Fractional response variables; censored distributions; Tobit models; limited dependent variables; Loss Given Default.

1. INTRODUCTION

Proportions or fractions are of considerable interest in economics as well as other sciences. They are usually bounded by 0 and 1 (or 100%). Often, such quantities show a substantial probability for adopting one or both of the boundary values. Such variables have been termed "fractional response variables" by Papke and Woolridge (1996). In a recent survey paper on modeling fractional data, Ramalho et al. (2011) list pension plan participation rates, firm market share, proportion of debt in the financing mix of firms, fraction of land area allocated to agriculture, and proportion of exports in total sales as examples. Another example is illustrated in Papke and Woolridge (2008), where test pass rates are analyzed.

In insurance, losses are frequently restricted to be positive and below an upper bound defined by a contract. We analyze a Loss Given Default dataset from an insurance category called "surety". In this example, claims cannot exceed a prespecified insured maximum, i.e., the ratio of loss over maximum is bounded by 1. On the other hand, for several reasons, the claims often do not lead to ultimate losses. The interest is in relating the distribution of this variable to a set of explanatory variables by a regression model.

Astin Bulletin 41(2), 673-710. doi: 10.2143/AST.41.2.2136992

For a fractional response variable Y, an important type of models focuses on the conditional mean $E[Y|\boldsymbol{x}]$ given a vector of covariates $\boldsymbol{x}$. A popular choice is to use the logistic function as a link function between a linear predictor and $E[Y|\boldsymbol{x}]$, but other cumulative distribution functions can also be used. Another semiparametric approach relies on assumptions about quantiles (see, e.g., Powell (1984), Khan and Powell (2001) or Chen and Khan (2001)). Whereas these approaches are sufficient for the purpose of many studies, in other cases, other aspects of the distribution of Y given $\boldsymbol{x}$, like upper quantiles or probabilities of attaining the limits, are of interest, as is the case in our application. In that case, parametric models are advantageous. On the other hand, since semiparametric models rely on less assumptions, they have the advantage that they are less prone to misspecification.

When there is a non-zero probability that the boundary values are attained, it is natural to use models based on censored random variables. These models are used in different fields of application. In economics, analyzing household expenditure on durable goods, Tobin (1958) first introduced such a model which later was coined Tobit model by Goldberger (1964). In climate science, precipitation can be modelled using censored distributions (see, e.g., Bardossy and Plate (1992) or Sanso and Guenni (2004)).

The Tobit model describes the distribution of Y given $\boldsymbol{x}$ as a censored normal with expectation $\mu = \boldsymbol{x}'\boldsymbol{\beta}$. It is therefore often perceived as a model for censored data, which it is in the detection limit case. However, it is perfectly adequate to use the censored normal distribution as a probability model in situations where no actual censoring occurs and the zeros are genuine values of the response, as is the case for the original application of Tobin (1958). The use of censored distributions is then a device to obtain a tractable model even though the data is not actually censored.

To support our thinking about the situation to be modeled, it is often helpful to attach the idea of a "potential" to a latent, uncensored response variable Y^*, of which Y is the censored version. In the case of precipitation, this potential measures a tendency for rain which may move from zero to way below, indicating that the weather develops from cloudy to very dry. For the standardized losses in our insurance example, the latent variables can be thought of as a loss potential. We note that Wooldridge (2010) calls models for variables that have a discrete and a continuous part, without actual censoring occurring, corner solution models. In Wooldridge (2002, Chapter 16), it is stated that an additional advantage of using a parametric distribution for modeling corner solution outcomes is that estimates of quantities such as $E[Y|\boldsymbol{x}]$ are efficient.

The Tobit model is easily adjusted to the case of an additional upper limit for Y (Rosett and Nelson (1975)) and thus to fractional data, and the generalization to replacing the normal distribution by any other suitable family is conceptually straightforward. In fact, when using censored distributions, one can model all quantities of interest, such as the mean, quantiles, and probabilities of attaining limits, together. We will focus on this approach in the following, using a shifted gamma distribution instead of the normal. The gamma distribution is a flexible distribution that is popular in insurance, and it will be shown to fit our data well (see, e.g., Figure 2).

Models based on a single random variable, such as the Tobit model or the censored gamma model, have the advantage of having a parsimonious parameterization entailing easier and more consistent interpretation. However, there are situations in which the frequencies of the limits do not follow this parsimonious description. We therefore also introduce two extensions of the model. For instance, in our example there may even be administrative reasons for an excessive number of zero losses, due to incentives to place a claim with little justification. Such preventive filing may result in a large number of "additional zeros". This idea suggests a mixture model, consisting of a censored part, as introduced above, and a model for the additional zeros.

An other approach to tackle this problem is called two-part models by Ramalho et al. (2011). These are extensions of the models of Papke and Wooldridge (1996). Here, a first model describes the occurrence of boundary values. Then, the continuous part can be modeled, for instance, by using the beta distribution (see Paolino (2001) and Ferrari and Cribari-Neto (2004)). Ramalho and da Silva (2009) and Cook et al. (2008) present empirical applications of two-part fractional response models. We also introduce an alternative extension of the censored gamma model based on this two-part modeling idea. Here, the probabilities of attaining the boundary value(s) are modelled separately from the continuous part in between them.

The rest of the paper is organized as follows. In Section 2, we introduce the censored gamma model, show how it can be interpreted, and derive an estimation procedure for it. In Section 3, two possible generalizations are presented. In Section 4, we illustrate an application of the models to the dataset mentioned above.

2. The Censored Gamma Model

In order to establish ideas, consider the Tobit model in its two sided version as developed by Rosett and Nelson (1975). It is assumed that there exists a latent variable Y^* which is, conditional on some covariates $\boldsymbol{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$, normally distributed. This variable is observed only if it lies in the interval $[0, 1]$. Otherwise, we observe 0 or 1, depending on whether the latent variable is smaller than 0 or greater than 1, respectively. If Y denotes the observed variable, this can be expressed as

$$Y^* \,|\, \boldsymbol{x} \sim \mathcal{N}(\mu, \sigma^2) \tag{1}$$

and

$$\begin{aligned} Y &= 0, \text{ if } Y^* \leq 0, \\ &= Y^*, \text{ if } 0 < Y^* < 1, \\ &= 1, \text{ if } Y^* \geq 1. \end{aligned} \tag{2}$$

Furthermore, the expectation μ of the latent variable Y^* is related to the covariates $\boldsymbol{x}$ through

$$\mu = \boldsymbol{x}'\boldsymbol{\beta}, \;\; \boldsymbol{\beta} \in \mathbb{R}^p.$$

For more details, e.g., on inference, we refer to Maddala (1983), Chapter 6, and Amemiya (1985), Chapter 10. Furthermore, Breen (1996) and Long (1997) give overviews of models for limited dependent variables.

Clearly, the assumption of a normal distribution for Y^* is not adequate for all data. It is well known that the Tobit model is sensitive to distributional assumptions (see, e.g., Arabmazar and Schmidt (1982) or Maddala and Nelson (1975)). A natural alternative is to replace the normal distribution by another one. We choose a shifted gamma distribution since it is a flexible distribution that is applied in many areas, especially in insurance. Further, it provides a good fit to the dataset of insurance claims mentioned above. This choice relies on distributional assumptions which have to be checked when applying the model to data.

To avoid unnecessary inflation of notation, we let the boundaries of the observed variable be 0 and 1. The model is easily generalized for variables whose range of values is any interval $[y_l, y_u]$ with $y_l < y_u$, though. This might be done either by first applying a linear transformation to the respective variable or by reformulating the model. The case where the observations are only bounded from below is included by letting $y_u \to \infty$.

2.1. The Model

Generalizing the Tobit model specified in (1) and (2), it is assumed that there exists a latent variable Y^* which has, conditional on $\boldsymbol{x}$, a distribution with density $f^*_{\boldsymbol{\theta}^*}(y^*)$ and cumulative distribution function $F^*_{\boldsymbol{\theta}^*}(y^*)$, $\boldsymbol{\theta}^*$ being a vector of parameters. The observed dependent variable Y then depends on the latent variable as in (2).

It follows that the distribution of such a censored variable Y can be characterized by

$$\begin{aligned} P[Y = 0] &= F^*_{\boldsymbol{\theta}^*}(0), \\ P[Y \in (y, y + dy)] &= f^*_{\boldsymbol{\theta}^*}(y)\,dy, \;\; 0 < y < 1, \\ P[Y = 1] &= 1 - F^*_{\boldsymbol{\theta}^*}(1). \end{aligned} \tag{3}$$

Consequently, the density of the observed variable Y can be written as

$$f_{\boldsymbol{\theta}^*}(y) = F^*_{\boldsymbol{\theta}^*}(0)\delta_0(y) + f^*_{\boldsymbol{\theta}^*}(y)\mathbf{1}_{\{0<y<1\}}(y) + (1 - F^*_{\boldsymbol{\theta}^*}(1))\delta_1(y), \;\; 0 \le y \le 1, \tag{4}$$

where $\delta_0(y)$ and $\delta_1(y)$ are Dirac measures and where $\mathbf{1}_{\{0<y<1\}}(y)$ denotes the indicator function equaling 1 if $0 < y < 1$ and 0 otherwise.

In order to extend the model to the regression case, we relate the distribution of Y^* to the covariates $\boldsymbol{x}$. This is done by assuming that the main parameter ϑ of the distribution of Y^*, which might be the mean or a scale parameter, is related through a link function g to the covariates,

$$g(\vartheta) = \boldsymbol{x}'\boldsymbol{\beta}. \tag{5}$$

In the following, we will focus on the case where the distribution of Y^* is specified as a gamma distribution with a shifted origin. The density and the distribution function of a gamma distributed variable with shape parameter α and scale parameter ϑ are denoted by $g_{\alpha,\vartheta}(y)$ and $G_{\alpha,\vartheta}(y)$, respectively. The density of a shifted gamma distribution is then

$$g_{\alpha,\vartheta}(y^* + \xi) = \frac{1}{\vartheta^\alpha \Gamma(\alpha)} (y^* + \xi)^{\alpha-1} e^{-(y^*+\xi)/\vartheta}, \quad y^* > -\xi,$$

where ξ, ϑ, $\alpha > 0$, and its distribution function is $G_{\alpha,\vartheta}(y^* + \xi)$.

The density of the observed Y can be expressed as

$$\begin{aligned} f_{\alpha,\vartheta,\xi}(y) = {} & G_{\alpha,\vartheta}(\xi)\delta_0(y) + g_{\alpha,\vartheta}(y+\xi)\mathbf{1}_{\{0<y<1\}}(y) \\ & + (1 - G_{\alpha,\vartheta}(1+\xi))\,\delta_1(y), \ \ 0 \le y \le 1. \end{aligned} \tag{6}$$

The use of a gamma distribution with a shifted origin, instead of a standard gamma distribution, is motivated by the fact that the lower censoring occurs at zero. In this case, the shift ξ is needed to obtain a positive probability of $Y = 0$.

For the regression case, we assume that the scale parameter ϑ is related to the covariates via the logarithmic link function

$$\log(\vartheta) = \boldsymbol{x}'\boldsymbol{\beta}. \tag{7}$$

Henceforth and if not otherwise stated, we assume that Y^* (and Y) follow a (censored) shifted gamma distribution. We will refer to this model as the “censored gamma model”.

Note that if no censoring occurred and ξ was set to zero, the censored gamma model would be a generalized linear model (McCullagh and Nelder (1983)) for a gamma distributed variable with a logarithmic link function.

2.2. Interpretation

If the focus lies on the latent response variable Y^*, the interpretation is straightforward. Since

$$E[Y^*|\boldsymbol{x}] = \alpha\vartheta - \xi, \tag{8}$$

the marginal effect of a continuous predictor x_j on $E[Y^*|\boldsymbol{x}]$ is

$$\frac{\partial E[Y^*|\boldsymbol{x}]}{\partial x_j} = \beta_j \alpha \vartheta. \tag{9}$$

On the other hand, one might be primarily interested in the observed variable Y, rather than the latent variable Y^*. Its mean and corresponding marginal effects are calculated in the following lemma.

Lemma 2.1. *The following holds true.*

$$\begin{aligned} E[Y|\boldsymbol{x}] = {} & \alpha\vartheta\big(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)\big) \\ & + (1+\xi)\big(1 - G_{\alpha,\vartheta}(1+\xi)\big) - \xi\big(1 - G_{\alpha,\vartheta}(\xi)\big), \end{aligned} \tag{10}$$

and for a continuous covariate x_j,

$$\frac{\partial E[Y|\boldsymbol{x}]}{\partial x_j} = \beta_j \alpha\vartheta(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)). \tag{11}$$

The derivation of these two equations is shown in Appendix A.

We note that the marginal effect of x_j on $E[Y|\boldsymbol{x}]$ is a scaled version of the effect on $E[Y^*|\boldsymbol{x}]$, with the scaling factor depending nonlinearly on the covariates.

If the interest lies on, say, the probability of Y being zero, $P[Y=0] = G_{\alpha,\vartheta}(\xi)$, one can also calculate partial effects on this quantity. For a continuous x_j, using similar ideas as in the proof of the above lemma, it is easily shown that

$$\begin{aligned} \frac{\partial P[Y=0|\boldsymbol{x}]}{\partial x_j} &= \frac{\partial G_{\alpha,\vartheta}(\xi)}{\partial x_j} \\ &= -\beta_j \xi g_{\alpha,\vartheta}(\xi). \end{aligned} \tag{12}$$

Finally, one can also consider quantiles. The quantile function $F^{\leftarrow}_{\alpha,\vartheta,\xi}(q)$, for $q \in [0,1]$, of Y is given by

$$\begin{aligned} F^{\leftarrow}_{\alpha,\vartheta,\xi}(q) &= 0, && \text{if } 0 \le q \le G_{\alpha,\vartheta}(\xi), \\ &= \vartheta G^{-1}_{\alpha,1}(q) - \xi, && \text{if } G_{\alpha,\vartheta}(\xi) < q < G_{\alpha,\vartheta}(1+\xi), \\ &= 1, && \text{if } G_{\alpha,\vartheta}(1+\xi) \le q \le 1. \end{aligned} \tag{13}$$

The partial effect of a continuous covariate x_j on the q-quantile $F^{\leftarrow}_{\alpha,\vartheta,\xi}(q)$ is therefore

$$\begin{aligned} \frac{\partial F^{\leftarrow}_{\alpha,\vartheta,\xi}(q)}{\partial x_j} &= 0, && \text{if } 0 < q < G_{\alpha,\vartheta}(\xi), \\ &= \beta_j \vartheta G^{-1}_{\alpha,1}(q), && \text{if } G_{\alpha,\vartheta}(\xi) < q < G_{\alpha,\vartheta}(1+\xi), \\ &= 0, && \text{if } G_{\alpha,\vartheta}(1+\xi) < q < 1. \end{aligned} \tag{14}$$

Note that for the cases $q = G_{\alpha,\vartheta}(\xi)$ and $q = G_{\alpha,\vartheta}(1+\xi)$, the function $F^{\leftarrow}_{\alpha,\vartheta,\xi}(q)$ is not differentiable with respect to x_j and, consequently, partial effects cannot be calculated.

2.3. Estimation

In this section, it is shown how to perform maximum likelihood estimation for the censored gamma model using a Newton-Raphson method known as Fisher's scoring algorithm (see, e.g., Fahrmeir and Tutz (2001)).

Denoting generically by $\boldsymbol{\theta}$ all parameters that are to be estimated and by $\ell(\boldsymbol{\theta})$ the log-likelihood, Fisher's scoring algorithm starts with an initial estimate $\hat{\boldsymbol{\theta}}^{(0)}$ and iteratively calculates (until convergence is achieved)

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \hat{\boldsymbol{\theta}}^{(k)} + I(\hat{\boldsymbol{\theta}}^{(k)})^{-1} s(\hat{\boldsymbol{\theta}}^{(k)}), \quad k = 0, 1, 2 \ldots,$$

where

$$s(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

denotes the score function, i.e., the first derivative of the log-likelihood, and

$$I(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[s(\boldsymbol{\theta})\, s(\boldsymbol{\theta})^T \right]$$

is the Fisher Information Matrix. How these two quantities are calculated for the censored gamma model is shown in the following.

First, we reparametrize the shape parameter α through

$$\alpha' = \log(\alpha) \tag{15}$$

to ensure that α attains only positive values. The parameters that are to be estimated, therefore, consist of $\boldsymbol{\theta} = (\alpha', \boldsymbol{\beta}, \xi)$.

Assuming that we have independent data $y_1 \ldots, y_n$ with covariates $\boldsymbol{x}_1 \ldots, \boldsymbol{x}_n$, the log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}).$$

Lemma 2.2. *The following relations hold true.*

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \alpha'} = {} & \frac{\alpha}{G_{\alpha,\vartheta_i}(\xi)} \left(-\psi(\alpha)\, G_{\alpha,\vartheta_i}(\xi) + H^{(1)}_{\alpha}\left(0, \frac{\xi}{\vartheta_i}\right)\right) \mathbf{1}_{\{y_i = 0\}} \\ & + \alpha(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i + \xi))\, \mathbf{1}_{\{0 < y_i < 1\}} \\ & - \frac{\alpha}{1 - G_{\alpha,\vartheta_i}(1+\xi)} \left(-\psi(\alpha)\, G_{\alpha,\vartheta_i}(1+\xi) + H^{(1)}_{\alpha}\left(0, \frac{1+\xi}{\vartheta_i}\right)\right) \mathbf{1}_{\{y_i = 1\}}, \end{aligned} \tag{16}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_k} = -x_{ik}\xi \frac{g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)} \mathbf{1}_{\{y_i=0\}} + x_{ik}\left(-\alpha + \frac{y_i+\xi}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}} + x_{ik}(1+\xi)\frac{g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\mathbf{1}_{\{y_i=1\}}, \tag{17}$$

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \xi} = \frac{g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\mathbf{1}_{\{y_i=0\}} + \left(\frac{\alpha-1}{y_i+\xi} - \frac{1}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}} - \frac{g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\mathbf{1}_{\{y_i=1\}}, \tag{18}$$

where

$$\psi(\alpha) = \frac{d\log(\Gamma(\alpha))}{d\alpha}$$

denotes the digamma function (see Abramowitz and Stegun (1964)) and the functions $H_\alpha^{(1)}$ and $H_\alpha^{(2)}$ *are defined as*[1]

$$H_\alpha^{(1)}(l,u) := \frac{1}{\Gamma(\alpha)}\int_l^u \log(y)y^{\alpha-1}\exp(-y)\,dy \tag{19}$$

and

$$H_\alpha^{(2)}(l,u) := \frac{1}{\Gamma(\alpha)}\int_l^u \log(y)^2 y^{\alpha-1}\exp(-y)\,dy. \tag{20}$$

The derivation of the scoring functions is shown in the following.

At first, we infer from (3) that the likelihood function of an interval censored gamma distribution can be written as

$$L_y(\alpha,\vartheta,\xi) = G_{\alpha,\vartheta}(\xi)\mathbf{1}_{\{y=0\}} + g_{\alpha,\vartheta}(y+\xi)\mathbf{1}_{\{0<y<1\}} + (1-G_{\alpha,\vartheta}(1+\xi))\mathbf{1}_{\{y=1\}} \tag{21}$$

which is equivalent to writing

$$L_y(\alpha,\vartheta,\xi) = G_{\alpha,\vartheta}(\xi)^{\mathbf{1}_{\{y=0\}}}\cdot g_{\alpha,\vartheta}(y+\xi)^{\mathbf{1}_{\{0<y<1\}}}\cdot(1-G_{\alpha,\vartheta}(1+\xi))^{\mathbf{1}_{\{y=1\}}}. \tag{22}$$

It follows that we can write the log-likelihood function $\ell_i(\boldsymbol{\theta})$ of an observation y_i as

$$\ell_i(\boldsymbol{\theta}) = \log(G_{\alpha,\vartheta_i}(\xi))\mathbf{1}_{\{y_i=0\}} + \log(g_{\alpha,\vartheta_i}(y_i+\xi))\mathbf{1}_{\{0<y_i<1\}} + \log(1-G_{\alpha,\vartheta_i}(1+\xi))\mathbf{1}_{\{y_i=1\}}$$

[1] We note that the functions $H_\alpha^{(1)}(l,u)$ and $H_\alpha^{(2)}(l,u)$ can be calculated using numerical integration. In our application, we did this by adaptive quadrature using the QUADPACK routines 'dqags' and 'dqagi' (Piessens et al. (1983)) available from Netlib.

$$
\begin{aligned}
&= \log(G_{\alpha,\vartheta_i}(\xi))\,\mathbf{1}_{\{y_i=0\}} \\
&\quad + \left(-\alpha\log(\vartheta_i) - \log(\Gamma(\alpha)) + (\alpha-1)\log(y_i+\xi) - \frac{y_i+\xi}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}} \\
&\quad + \log(1 - G_{\alpha,\vartheta_i}(1+\xi))\,\mathbf{1}_{\{y_i=1\}},
\end{aligned}
$$

where

$$
\vartheta_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \quad \text{and} \quad \alpha = \exp(\alpha').
$$

The derivative of ℓ_i with respect to the parameter α' in (16) is then calculated using the following identity.

$$
\begin{aligned}
\frac{\partial G_{\alpha,\vartheta}(\xi)}{\partial\alpha} &= \frac{\partial G_{\alpha,1}\left(\frac{\xi}{\vartheta}\right)}{\partial\alpha} \\
&= \frac{\partial}{\partial\alpha}\left(\frac{1}{\Gamma(\alpha)}\int_0^{\xi/\vartheta} y^{\alpha-1}\exp(-y)\,dy\right) \\
&= \frac{-\Gamma'(\alpha)}{\Gamma(\alpha)^2}\int_0^{\xi/\vartheta} y^{\alpha-1}\exp(-y)\,dy + \frac{1}{\Gamma(\alpha)}\int_0^{\xi/\vartheta}\log(y)\,y^{\alpha-1}\exp(-y)\,dy \\
&= -\psi(\alpha)\,G_{\alpha,\vartheta}(\xi) + H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta}\right). \qquad (23)
\end{aligned}
$$

Next, using

$$
\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\beta_k} = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\vartheta_i}\frac{\partial\vartheta_i}{\partial\beta_k} = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\vartheta_i}\vartheta_i x_{ik}
$$

and (41), differentiating $\ell_i(\boldsymbol{\theta})$ with respect to β_k gives the result in (17). The calculation of the derivative with respect to ξ in (18) is straightforward.

For the Fisher-scoring algorithm and for asymptotic inference, we calculate the Fisher Information Matrix

$$
I(\boldsymbol{\theta})_{k,l} = E_\theta\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial \ell(\boldsymbol{\theta})}{\partial\theta_l}\right], \quad 1 \le k, l \le 2+p.
$$

Because of the independence of the observations, this can be written as

$$
\begin{aligned}
I(\boldsymbol{\theta})_{k,l} &= E_\theta\left[\left(\sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\theta_k}\right)\left(\sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\theta_l}\right)\right] \\
&= \sum_{i=1}^n E_\theta\left[\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\theta_k}\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial\theta_l}\right].
\end{aligned}
$$

The specific calculations of the entries $E_\theta\left[\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \theta_l}\right]$ are shown in Section S.1 in the supplementary material.

As mentioned before, the Fisher Information Matrix $I(\boldsymbol{\theta})$ is used in the Fisher-scoring algorithm for fitting the model and for asymptotic inference, in particular to estimate standard errors of the coefficients $\boldsymbol{\beta}$.

3. Two Extensions of the Model

A salient feature of the model defined in (3) and of the Tobit model is the assumption that the same parameters govern both the behaviour of the uncensored values as well as the probabilities of being censored from below or above.

In order to relax this assumption, various extensions have been proposed. Sample selection models, first introduced by Heckman (1976), are one approach. Cragg (1971) came forward with another proposal relaxing the aforementioned assumption of one set of parameters governing the entire model.

For count data, similar problems can arise: there may be more zeros than expected by a simple model, which would otherwise fit well. Basically, two different kinds of solutions have been put forward there.

Aitchison (1955) first proposed to model the zeros and the values bigger than zero separately. Mullahy (1986) used a mixture consisting of a distribution for the whole range of data, including zeros, and a point mass at zero to capture extra zeros. These two types of models have been extensively applied in various areas of research including manufacturing defects (Lambert (1992)), patent applications (Crepon and Duguet (1997)), road safety (Miaou (1994)), species abundance (Welsh et al. (1996)), medical consultations (Gurmu (1997)), use of recreational facilities (Gurmu and Trivedi (1996); Shonkwiler and Shaw (1996)), and sexual behaviour (Heilbron (1994)). Ridout et al. (1998) give an overview of these models.

Our two extensions are based on similar ideas. The main difference is the way in which the zeros are modeled. In the first extension, the zeros and the non-zero values are modeled separately assuming that the mechanisms that govern the probability of Y being zero and the non-zero part are different. In the second extension, the zeros are modelled as a mixture of two mechanisms. One is responsible for artificial or extra zeros whereas the other part is the censored gamma model introduced in Section 2.

3.1. The Two-tiered Gamma Model

Inspired by the approach of Cragg (1971), we extend the model in (3) by allowing for two different sets of parameters, one governing the probability of Y being zero, and the other the behaviour for $0 < Y \leq 1$.

Alternatively, the model could also be extended by allowing for a different set of parameters governing the probability of Y being one. The extension presented here, which we will call two-tiered gamma model, is mainly motivated by the presumption that zeros are generated by another mechanism than the one that governs the rest of the data. We remark that the extension to a "three-tiered" model including a different set of parameters for governing the probability of Y being one is straightforward.

More specifically, in the two-tiered gamma model, it is assumed that there exist two latent variables

$$Y_1^* \sim G_{\alpha,\tilde{\vartheta}}(y_1^* + \xi), \;\text{ with }\; \tilde{\vartheta} = \exp(\boldsymbol{x}'\boldsymbol{\gamma}), \;\; \boldsymbol{\gamma} \in \mathbb{R}^p$$

and

$$Y_2^* \sim G_{\alpha,\vartheta}(y_2^* + \xi) \text{ truncated at } 0, \;\text{ with }\; \vartheta = \exp(\boldsymbol{x}'\boldsymbol{\beta}), \;\; \boldsymbol{\beta} \in \mathbb{R}^p.$$

The first latent variable Y_1^* is again following a shifted gamma distribution, whereas the second variable Y_2^* has shifted gamma distribution that is lower truncated at zero. These two latent variables are then related to Y through

$$\begin{aligned} Y &= 0 \quad \text{if } \; Y_1^* \leq 0, \\ &= Y_2^* \quad \text{if } \quad 0 < Y_1^* \text{ and } Y_2^* < 1, \\ &= 1 \quad \text{if } \quad 0 < Y_1^* \text{ and } 1 \leq Y_2^*. \end{aligned}$$

In other words, the two-tiered gamma model first decides whether Y is zero or not. This is modeled in the style of a probit model, using, however, a cumulative gamma distribution function instead of a normal one. It is then assumed that, conditional on $Y > 0$, $0 < Y \leq 1$ has a lower truncated and upper censored gamma distribution.

The distribution of Y can then be characterized as follows.

$$\begin{aligned} P[Y = 0] &= G_{\alpha,\tilde{\vartheta}}(\xi), \\ P[Y \in (y, y + dy)] &= g_{\alpha,\vartheta}(y + \xi)\, \frac{1 - G_{\alpha,\tilde{\vartheta}}(\xi)}{1 - G_{\alpha,\vartheta}(\xi)}\, dy, \;\; 0 < y < 1, \\ P[Y = 1] &= (1 - G_{\alpha,\vartheta}(1 + \xi))\, \frac{1 - G_{\alpha,\tilde{\vartheta}}(\xi)}{1 - G_{\alpha,\vartheta}(\xi)}, \end{aligned} \tag{24}$$

with

$$\vartheta = \exp(\boldsymbol{x}'\boldsymbol{\beta}), \;\; \tilde{\vartheta} = \exp(\boldsymbol{x}'\boldsymbol{\gamma}), \; \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{R}^p, \;\; \alpha, \xi > 0.$$

Again, $g_{\alpha,\vartheta}(y)$ denotes the density of a Gamma(α, ϑ) distributed variable and $G_{\alpha,\vartheta}(y)$ is the corresponding distribution function.

We remark that the distributions in both parts of the two-tiered model, i.e., the part modeling the probability of Y being zero and the part governing the behaviour of $0 < Y \leq 1$, are assumed to have the same shape parameter α and the same location parameter ξ. Consequently, if $\boldsymbol{\beta} = \boldsymbol{\gamma}$, or $\vartheta = \tilde{\vartheta}$, the two-tiered gamma model presented here and the aforementioned censored gamma model coincide, which means that these two models are nested. This is convenient for model comparison since it allows to use a likelihood ratio test to compare the two models.

3.2. Estimation of the Two-tiered Gamma Model

Having in mind that the censored gamma model is nested in the two-tiered gamma model, we restrict ourselves to estimating the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of the two linear predictors using Fisher's scoring algorithm. The shape parameter α and the location parameter ξ could be estimated via numerical optimization in an outer loop with starting values obtained from first fitting a censored gamma model.

With $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, the log-likelihood function of the model can be written as $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$ with

$$\begin{aligned}\ell_i(\boldsymbol{\theta}) = {} & \log(G_{\alpha,\tilde{\vartheta}_i}(\xi))\mathbf{1}_{\{y_i=0\}} \\ & + (\log(g_{\alpha,\vartheta_i}(y_i+\xi)) + \log(1-G_{\alpha,\tilde{\vartheta}_i}(\xi)) - \log(1-G_{\alpha,\vartheta_i}(\xi)))\mathbf{1}_{\{0<y_i<1\}} \\ & + (\log(1-G_{\alpha,\vartheta_i}(1+\xi) + \log(1-G_{\alpha,\tilde{\vartheta}_i}(\xi)) - \log(1-G_{\alpha,\vartheta_i}(\xi)))\mathbf{1}_{\{y_i=1\}},\end{aligned}$$

where

$$\vartheta_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}), \;\; \tilde{\vartheta}_i = \exp(\boldsymbol{x}_i'\boldsymbol{\gamma}).$$

The score functions are

$$\begin{aligned}\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \beta_k} = {} & x_{ik}\left(\frac{y_i+\xi}{\vartheta_i} - \alpha - \frac{\xi \cdot g_{\alpha,\vartheta_i}(\xi)}{1-G_{\alpha,\vartheta_i}(\xi)}\right)\mathbf{1}_{\{0<y_i<1\}} \\ & + x_{ik}\cdot\left(\frac{(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)} - \frac{\xi \cdot g_{\alpha,\vartheta_i}(\xi)}{1-G_{\alpha,\vartheta_i}(\xi)}\right)\mathbf{1}_{\{y_i=1\}}\end{aligned} \tag{25}$$

and

$$\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \gamma_k} = -x_{ik}\frac{\xi \cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{G_{\alpha,\tilde{\vartheta}_i}(\xi)}\mathbf{1}_{\{y_i=0\}} + x_{ik}\frac{\xi \cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{1-G_{\alpha,\tilde{\vartheta}_i}(\xi)}\cdot\left(\mathbf{1}_{\{0<y_i<1\}} + \mathbf{1}_{\{y_i=1\}}\right). \tag{26}$$

The entries of the Fisher Information Matrix $I(\boldsymbol{\theta})$ are presented in Appendix S.2.

3.3. The Zero-Inflated Gamma Model

The extension presented in this section is motivated by the following idea. Assume that our quantity of interest follows indeed a censored, shifted gamma distribution. However, additional, artificial zeros occur by some other mechanism and thus there are more zeros than expected. Deaton and Irish (1984) used such an extension of the Tobit model for modeling expenditures in household budgets. Recently, a zero-inflated model for censored continuous data has also been presented by Couturier and Victoria-Feser (2010).

These additional zeros are now allowed to follow their own model, in contrast to the two-tiered model where all zeros were described together. This view may make sense in specific applications like insurance, where some of the claims that result in zero losses may be cases which were filed in order not to miss a formal deadline or for similar artificial reasons.

In the zero-inflated model, the existence of two latent variables is again assumed,

$$Y_1^* \sim N(-\mu, 1) \quad \text{and} \quad Y_2^* \sim G_{\alpha,\vartheta}(y_2^* + \xi)$$

with $\mu = \boldsymbol{x}'\boldsymbol{\gamma}$ and $\vartheta = \exp(\boldsymbol{x}'\boldsymbol{\beta})$.

We note that the censored gamma model is not nested in the zero-inflated model in the classical sense. However, the zero-inflated model coincides with the censored gamma model at the boundary of its parameter space, namely if $\mu \to -\infty$. For the reason of simplicity, we opt for the normal distribution. I.e., the extra zeros are model using a probit model. Alternatively, one could also use the logit distribution.

These two variables are then related to Y through

$$\begin{aligned} Y &= 0 \quad \text{if } Y_1^* \le 0, \quad \text{or if} \\ &\qquad\qquad 0 < Y_1^* \text{ and } Y_2^* \le 0, \\ &= Y_2^* \text{ if } \quad 0 < Y_1^* \text{ and } \quad 0 < Y_2^* < 1, \\ &= 1 \quad \text{if } \quad 0 < Y_1^* \text{ and } \qquad 1 \le Y_2^*. \end{aligned}$$

The variable Y_1^* first decides whether the observed response variable Y is zero, i.e., if $Y_1^* \le 0$ it follows that $Y = 0$. Next, conditional on $Y_1^* > 0$, Y is distributed according to a censored, shifted gamma distribution.

This means that the zeros are governed by two different components of the model. First, zeros can arise if Y_1^* is smaller than zero. And secondly, they can occur if, conditional on $Y_1^* > 0$, Y_2^* is smaller than zero. Metaphorically speaking, we add extra mass at zero to the censored gamma distribution, which can account for potential extra zeros. This approach allows us to distinguish structural and extra zeros.

Note that the main distinctive feature of this model, in contrast to the two-tiered model presented in the previous section, is that the distribution of the second tier of the model is lower censored instead of lower truncated.

As stated above, we choose to model the extra zeros using a probit model, i.e.,

$$p_0 := P[Y_1^* \le 0] = \Phi(\boldsymbol{x}'\boldsymbol{\gamma}), \; \boldsymbol{\gamma} \in \mathbb{R}^p. \tag{27}$$

Consequently, the distribution of Y can be characterized by

$$\begin{aligned} P[Y=0] &= p_0 + (1-p_0)\cdot G_{\alpha,\vartheta}(\xi), \\ P[Y\in(y,y+dy)] &= (1-p_0)\cdot g_{\alpha,\vartheta}(y+\xi)dy,\ \ 0<y<1, \\ P[Y=1] &= (1-p_0)\cdot(1-G_{\alpha,\vartheta}(1+\xi)), \end{aligned} \tag{28}$$

where

$$p_0 = \Phi(\boldsymbol{x}'\gamma),\ \ \vartheta = \exp(\boldsymbol{x}'\boldsymbol{\beta}),\ \ \gamma,\boldsymbol{\beta}\in\mathbb{R}^p,\ \ \alpha,\xi>0.$$

We note that the zero-inflated model reduces to the censored Gamma model in the limit $\mu\rightarrow-\infty$, i.e., at the boundary of the parameter space. This means that a straightforward likelihood ratio test for model selection does not apply here. In Section 4.2 in the application, we use a simulation based testing procedure to compare these two models.

3.4. Estimation of the Zero-Inflated Gamma Model

Since the EM (Dempster et al. (1977)) algorithm lends itself naturally when it comes to fitting mixtures of distributions and because calculations of scores and the Fisher Information Matrix would be overly complicated, we use the EM algorithm here.

The EM algorithms presented in the following finds the maximum likelihood estimators of the parameters $\boldsymbol{\theta}=(\alpha,\boldsymbol{\beta},\gamma)$. The location parameter ξ is fixed and assumed to be known. Again, ξ could be obtained from first fitting the censored gamma model or it could be estimated through numerical optimization in an outer loop. Alternatively, the values obtained from the EM Algorithm together with the estimated ξ from the censored gamma model can be used as starting values for generic optimization algorithms such as, for instance, quasi-Newton methods. We note that in some examples we observed convergence problems when using quasi-Newton methods without reasonable starting values.

With regard to the EM algorithm, we introduce two latent data variables Z and Y^*. For each i, Z_i indicates whether the observation belongs to the extra zero part of the model ($Z_i=0$) or to the censored gamma distribution ($Z_i=1$). The second missing data variable Y_i^* is for the censored gamma part of the model. It denotes the value of the underlying latent variable Y_i^* which then is censored at zero and one. The complete data $\boldsymbol{W}$ therefore consists of $(Z_1,Y_1^*)\ldots,(Z_n,Y_n^*)$.

Using this, the complete-data likelihood can be written as

$$L_{\boldsymbol{W}}(\boldsymbol{\theta}) = \prod_{i=1}^{n}\Big(\Phi(\boldsymbol{x}_i'\gamma)\Big)^{1-Z_i}\cdot\Big((1-\Phi(\boldsymbol{x}_i'\gamma))\cdot g_{\alpha,\vartheta_i}(Y_i^*+\xi)\Big)^{Z_i}, \tag{29}$$

where $\log(\vartheta_i)=\boldsymbol{x}_i'\boldsymbol{\beta}$ and $\boldsymbol{\theta}=(\alpha,\boldsymbol{\beta},\gamma)$, and the complete-data log-likelihood is

$$\begin{aligned}\ell_W(\boldsymbol{\theta}) &= \sum_{i=1}^{n} (1-Z_i)\log\Big(\Phi(\boldsymbol{x}_i'\boldsymbol{\gamma})\Big) + Z_i \log\Big((1-\Phi(\boldsymbol{x}_i'\boldsymbol{\gamma}))\cdot g_{\alpha,\vartheta_i}(Y_i^* + \xi)\Big)\\ &= \sum_{i=1}^{n} (1-Z_i)\log\Big(\Phi(\boldsymbol{x}_i'\boldsymbol{\gamma})\Big) + Z_i \log\Big(1-\Phi(\boldsymbol{x}_i'\boldsymbol{\gamma})\Big) \\ &\quad + \sum_{i=1}^{n} Z_i\left(-\alpha\log(\vartheta_i) - \log(\Gamma(\alpha)) + (\alpha-1)\log(Y_i^*+\xi) - \frac{Y_i^*+\xi}{\vartheta_i}\right).\end{aligned} \tag{30}$$

The EM algorithm produces a sequence of estimates $\{\boldsymbol{\theta}^{(t)},\ t = 0, 1, 2, \ldots\}$ by alternatively applying two steps:

E-step. Compute the expected value of the log-likelihood, with respect to the conditional distribution of $\boldsymbol{W}$ given $\boldsymbol{y}$ under the current estimate of the parameters $\boldsymbol{\theta}^{(t)}$:

$$Q^{(t+1)}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}^{(t)}}[\ell_W(\boldsymbol{\theta})\,|\,\boldsymbol{y}].$$

M-step. Update the parameter estimated according to:

$$\boldsymbol{\theta}^{(t+1)} = \operatorname{argmax}_{\boldsymbol{\theta}}\, Q^{(t+1)}(\boldsymbol{\theta}).$$

From (30), we infer that in the E-step three different expectations have to be calculated: $E_{\boldsymbol{\theta}}^{(t)}[Z_i|\boldsymbol{y}]$, $E_{\boldsymbol{\theta}}^{(t)}[Y_i^*+\xi|\boldsymbol{y}]$, and $E_{\boldsymbol{\theta}}^{(t)}[\log(Y_i^*+\xi)|\boldsymbol{y}]$. For the sake of notational brevity, we introduce the following two abbreviations:

$$A_i^{(t)} = \Phi(\boldsymbol{x}_i'\boldsymbol{\gamma}^{(t)})$$

and

$$B_i^{(t)}(\xi) = G_{\alpha^{(t)},\,\vartheta_i^{(t)}}(\xi).$$

The three expectations are then calculated as follows:

$$E_{\boldsymbol{\theta}^{(t)}}[Z_i\,|\,\boldsymbol{y}] = \begin{cases} \dfrac{(1-A_i^{(t)})\cdot B_i^{(t)}(\xi)}{A_i^{(t)} + (1-A_i^{(t)})\cdot B_i^{(t)}(\xi)} & \text{if } y_i = 0,\\ 0 & \text{if } y_i > 0, \end{cases} \tag{31}$$

$$E_{\boldsymbol{\theta}^{(t)}}[Y_i^*+\xi\,|\,\boldsymbol{y}] = \begin{cases} \alpha^{(t)}\vartheta_i^{(t)}\,\dfrac{G_{\alpha^{(t)}+1,\,\vartheta_i^{(t)}}(\xi)}{B_i^{(t)}(\xi)} & \text{if } y_i = 0,\\ y_i + \xi & \text{if } 0 < y_i < 1,\\ \alpha^{(t)}\vartheta_i^{(t)}\,\dfrac{\Big(1 - G_{\alpha^{(t)}+1,\,\vartheta_i^{(t)}}(1+\xi)\Big)}{\Big(1 - B_i^{(t)}(1+\xi)\Big)} & \text{if } y_i = 1, \end{cases} \tag{32}$$

and

$$E_{\theta^{(t)}}[\log(Y_i^* + \xi) \mid \boldsymbol{y}] = \begin{cases} \log(\vartheta_i^{(t)}) + \dfrac{H^{(1)}_{\alpha^{(t)}}\left(0, \frac{\xi}{\vartheta_i^{(t)}}\right)}{B_i^{(t)}(\xi)} & \text{if } y_i = 0, \\ \log(y_i + \xi) & \text{if } 0 < y_i < 1, \\ \log(\vartheta_i^{(t)}) + \dfrac{H^{(1)}_{\alpha^{(t)}}\left(\frac{1+\xi}{\vartheta_i^{(t)}}, \infty\right)}{1 - B_i^{(t)}(1+\xi)} & \text{if } y_i = 1. \end{cases} \quad (33)$$

Concerning the M-step, we note that the log-likelihood in (30) splits into two terms which can be maximized separately. The first term contains the parameters of the extra zero model part ($\boldsymbol{\gamma}$) and the other contains the parameters of the censored gamma distribution (α and $\boldsymbol{\beta}$).

4. An Application

4.1. Loss Given Default Data

We apply the models presented above to a dataset from insurance. A surety bond is a contractual agreement among three parties: the contractor who performs an obligation, the obligee who receives the obligation, and the surety provider, in our case the insurance company, who covers the risk that the contractor fails to fulfill the obligation.

The dataset consists of European surety bonds that resulted in a claim. The ultimate loss for these claims is called "Loss Given Default" (LGD). For each bond, the maximal amount that is covered by the insurance company, a quantity called "face value" (FV), is a priori determined. This allows us to standardize the LGD by dividing it by the face value, such that our variable of interest lies between 0 and 1

$$0 \leq \frac{LGD}{FV} \leq 1. \quad (34)$$

We have worked with the original dataset, but for confidentiality reasons the results presented here are obtained on the basis of a subsample of the original set. The subsample, consisting of more than 5000 bonds, is obtained by using a random selection mechanism, with selection probabilities that depend on certain characteristics of the respective bonds, so that the value of the average standardized loss LGD/FV is altered in order not to reveal the true average. As a consequence, the results presented in this paper are not the real ones but are close enough to reflect the major phenomena. We assure that the fit the models provide to the original data is at least as good as for the subsample.

The standardized losses are shown in Figure 1. Since the insurance company can often recover costs, observations with no ultimate loss at all are frequent. In fact, about 52% of all bonds in the subsample have no loss.

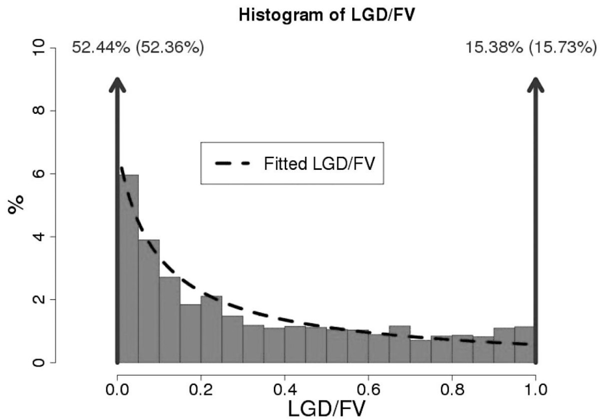

FIGURE 1: Histogram of LGD/FV and fitted censored gamma model with no covariates. The numbers above the blue arrows represent the percentage of LGD/FV's being exactly zero or one, respectively. In parentheses are the corresponding numbers as predicted by the censored gamma model. The dashed red line represents the fitted model.

On the other hand, there is a major proportion (15%) of bonds that have full loss, i.e., a LGD/FV equaling 1.

Apart from providing a probabilistic model for the surety LGD, the purpose is also to explore the relation of the losses to certain covariates which are shortly described in the following.

The relative default time (RDT) of a bond is the proportion of time that has passed at default since its issuance over the total life span of a bond. This

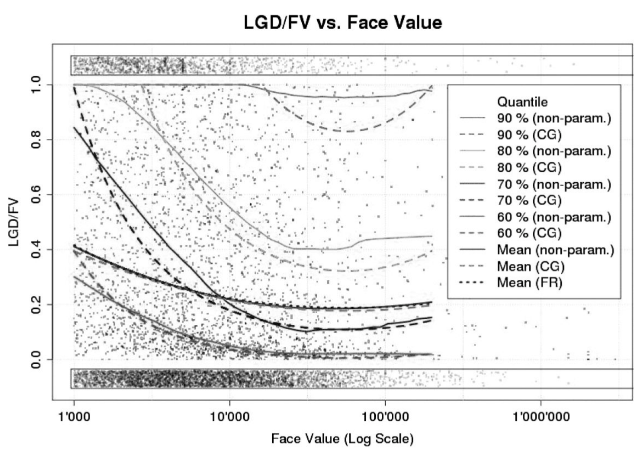

FIGURE 2: Scatter plot of face value (on a logarithmic scale) vs. LGD/FV. The jittered points in the bars below 0.0 and above 1.0 represent bonds with LGD/FV being exactly zero and one, respectively. The colored solid lines are non-parametrically fitted quantiles and mean. The dashed lines represent quantiles and mean of the fitted censored gamma (CG) model. The green dotted line represents the fitted conditional mean of the fractional response (FR) model. Logarithmic and squared logarithmic face value are taken as covariates.

quantity allows us to explore the time development of the losses from the issuing date to the end date (maturity). Experience and size are two categorical variables, each attaining three different levels, which represent the experience (low, mid, high) and the size (small, medium, large) of the contractor. There are three different types of surety bonds called maintenance, performance, and hybrid bonds. Hybrid bonds are bonds that are both maintenance and performance bonds. There is an additional category denoted "other bonds" for a small number of bonds of various other categories. Usually, European surety bonds do not cover the whole amount of an underlying contract but only a certain fraction. Information about his percentage is included as an additional covariate. In Table 2 in the supplementary material, we report summary statistics for the continuous covariates and relative frequencies for the categorical variables.

4.2. Results

We first estimate the censored gamma model of Section 2 with no covariates and illustrate its fit in Figure 1. The dashed red line represents the fitted model. The numbers in parentheses above the bars show the fitted probabilities of being zero and one. Apparently, the plain model with no covariates fits the data well. The observed and the modeled probabilities of being zero or one are very similar and the continuous part of the model accurately fits the histogram[2]. For comparison, we have also fitted the standard normal Tobit model in its two-sided version, as well as a corresponding model using a skewed t distribution (Azzalini and Capitanio (2003)). See the supplementary material for more details. Both models provide worse fits than the censored gamma model. A plot (Figure 4) illustrating the fits can be found in the supplementary material.

Next, we fit a model using only the face value, more specifically the logarithm and the squared logarithm of the face value, as covariate. We illustrate the fitted model in Figure 2. The colored continuous lines are non-parametrically fitted quantile (see Koenker (2005)) and mean curves (calculated using local polynomial regression, see Chambers and Hastie (1992), Chapter 8). The dashed lines represent the corresponding quantiles and mean of the fitted model calculated using the result in Lemma 2.1. We also fit the conditional mean model for fractional response (FR) of Papke and Wooldridge (1996). Here, fitting is done using quasi-maximum likelihood (see Gourieroux et al. (1984) for details) based on the Bernoulli log-likelihood function.

The non-parametrically fitted mean and the mean of the fitted censored gamma model are very close together. This indicates that the censored gamma model provides a good fit to the conditional mean. Moreover, the non-parametrically estimated quantiles and the quantiles from the fitted censored gamma model match well. I.e., the censored gamma model not only models the mean appropriately but the entire distribution. In addition, the fitted mean

[2] Due to the large number of observations, a chi-square goodness of fit test still shows significant deviations.

of fractional response model is very close the mean of the fitted censored gamma model. Again, we have also fitted the Tobit model and the skewed t version. Compared to the censored gamma model, both models provide worse fits (see Figure 5 in the supplementary material).

Finally, we fit the censored gamma and its two extensions, i.e., the two-tiered and the zero-inflated model including all covariates. For the two ordinal factorial variables experience and size, we use orthogonal polynomial contrasts. Concerning the categorical variable type, we use treatment contrasts with maintenance as baseline level. For the censored gamma model, we use the Fisher scoring algorithm presented above. In the case of the two-tiered and zero-inflated models, we use the algorithms presented in this paper to determine good starting values for quasi-Newton methods. Starting values for the parameters that are not estimated with these methods, i.e., the shape parameter α and the location parameter ξ, respectively, are obtained by taking the values from the ones in the fitted censored gamma model. We then estimate the two models using quasi-Newton methods. Concerning the censored gamma model, estimates of standard errors are calculated using the Fisher information. For the other two models, standard errors are obtained by numerically approximating the Fisher Information Matrix at the optimum.

The results are reported in Table 1. The log-likelihood of both the two-tiered and zero-inflated models are considerably higher than the one of the censored gamma model. This is also reflected in considerably smaller AIC values, the zero-inflated model having the lowest AIC. A likelihood ratio test clearly favors the two-tiered model over the censored gamma model. This is also true for the zero-inflated model. For the latter, the null hypothesis is on the boundary of the parameter space, and the usual asymptotics do not apply. We therefore use a simulated test instead. To be more specific, the distribution of the difference in log-likelihoods between the two models under the null hypothesis is characterized by 1000 simulated values. A sample from this distribution is generated by simulating data from the null hypothesis, i.e., from the estimated censored Gamma model, then fitting both models, and calculating the difference in the two log-likelihoods. The lowest simulated difference obtained out of the 1000 samples was about 28.6. We conclude that the observed difference of more than 200 is clearly significant. Next, for discriminating between the two extended models, we apply Vuong's test (Vuong (1989)). Since we know that the zero-inflated model does not reduce to the censored gamma model, it follows that we are not in the overlapping case. Thus, we can use the Vuong's non-nested hypothesis test. The test statistic has a value of -2.26 under the null hypothesis that both models are equally close to the true model. Thus, at a 5% level, the null hypothesis is rejected in favor of the zero-inflated model. This gives support to the idea that there are indeed extra zeros in the data. These extra zeros are interpreted as zero losses from claims that were filed for administrative reasons and not because there was a true default event. As before, we have also fitted the Tobit model and skewed t distribution model using all covariates. The Results are reported in Tables 3 and 4 in the supplementary material. In all cases, the gamma models have considerably lower

TABLE 1

FITTED CENSORED, TWO-TIERED, AND ZERO-INFLATED GAMMA MODELS INCLUDING ALL COVARIATES.
CODES FOR SIGNIFICANCE LEVELS: '***': $p < 0.001$, '**': $0.001 \leq p < 0.01$, '*': $0.01 \leq p < 0.05$, '.': $0.05 \leq p < 0.1$.

Covariate	Model	Censored		Two-Tiered				Zero-Inflated			
		Coef	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.
Intercept		3.9	0.34***	3.9	0.33***	−3.2	0.61***	4.1	0.35***	0.023	0.18
RDT	Lin	−0.17	0.10 ·	0.30	0.10**	−0.45	0.079***	0.29	0.10**	0.35	0.057***
	Quad	0.074	0.35	1.6	0.35***	−1.1	0.26***	1.6	0.35***	0.88	0.20***
Experience	Lin	−0.82	0.076***	−0.39	0.064***	−0.67	0.066***	−0.38	0.065***	0.42	0.037***
	Quad	0.12	0.051*	0.064	0.045	0.068	0.041 ·	0.059	0.046	−0.017	0.026
Size	Lin	0.56	0.32 ·	0.35	0.37	0.44	0.24 ·	0.34	0.38	−0.36	0.19 ·
	Quad	0.66	0.20**	−0.17	0.24	0.85	0.15***	−0.18	0.24	−0.68	0.12***
Face Value	Lin	−0.80	0.071***	−0.96	0.065***	−0.0048	0.050	−0.99	0.070***	−0.054	0.047
	Quad	0.50	0.068***	0.15	0.053**	0.49	0.064***	0.18	0.054**	−0.33	0.043***
Type	Hybrid	2.9	1.5 ·	2.0	1.2 ·	2.7	1.2*	1.7	1.1	−1.9	0.80*
	Performance	0.015	0.12	0.16	0.11	−0.12	0.099	0.17	0.11	0.12	0.070 ·
	Other	0.23	0.16	0.52	0.17**	−0.20	0.12	0.57	0.17**	0.19	0.095*
Ins. Frac.		1.2	0.56*	1.5	0.49**	−0.43	0.39	1.6	0.49***	0.40	0.28
		Value	Std. Err.		Value	Std. Err.			Value	Std. Err.	
Gamma Par.	$\log(\alpha)$	−1.5	0.050		−0.54	0.067			−0.57	0.073	
	$\log(\xi)$	−2.4	0.093		−4.5	0.47			−4.3	0.44	
Log-Likelihood		−7898.4		−7684.9				−7680.5			
AIC		15826.8		15425.9				15417.1			

AICs, and the corresponding differences in log-likelihood are always larger than 100, except when comparing the two-tiered gamma model with the two-tiered skewed t model where the differences is about 8 in favor of the gamma model. This means that Vuong's test favors the gamma model in all cases.

4.3. Interpretation of Results

Having come to the conclusion that the zero-inflated model provides the best fit to our data, we interpret the obtained results. Interpretation is not as straightforward as, for instance, in the basic censored gamma model case (see Section 2.2). In contrast to that, in the zero-inflated extension there are two linear predictors $\eta = \boldsymbol{x}'\boldsymbol{\beta}$ and $\mu = \boldsymbol{x}'\boldsymbol{\gamma}$. Partial effects on, say, the conditional mean therefore include both sets of coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. We will focus on $E[Y|\boldsymbol{x}]$ and $P[Y=0|\boldsymbol{x}]$ in the following. These two quantities and their corresponding partial effects are calculated in the following lemma.

Lemma 4.1. *For the zero-inflated model, the following relations hold true.*

$$E[Y\,|\,\boldsymbol{x}] = (1-\Phi(\mu))\,C^1_{\alpha,\vartheta,\xi} \tag{35}$$

where

$$\begin{aligned} C^1_{\alpha,\vartheta,\xi} = {} & \alpha\vartheta\big(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)\big), \\ & + (1+\xi)\big(1-G_{\alpha,\vartheta}(1+\xi)\big) - \xi\big(1-G_{\alpha,\vartheta}(\xi)\big), \end{aligned} \tag{36}$$

and

$$P[Y=0\,|\,\boldsymbol{x}] = \Phi(\mu) + (1-\Phi(\mu))\cdot G_{\alpha,\vartheta}(\xi). \tag{37}$$

For a continuous covariate x_j, we have

$$\frac{\partial E[Y|\boldsymbol{x}]}{\partial x_j} = \beta_j C^2_{\alpha,\vartheta,\xi}(1-\Phi(\mu)) - \gamma_j\,\phi(\mu)\,C^1_{\alpha,\vartheta,\xi}, \tag{38}$$

where

$$C^2_{\alpha,\vartheta,\xi} = \alpha\vartheta(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)), \tag{39}$$

and

$$\frac{\partial P[Y=0\,|\,\boldsymbol{x}]}{\partial x_j} = -\beta_j\xi g_{\alpha,\vartheta}(\xi)(1-\Phi(\mu)) + \gamma_j\phi(\mu)\big(1-G_{\alpha,\vartheta}(\xi)\big). \tag{40}$$

The lemma follows from (28) together with Lemma 2.1. We see that the partial effects contain $\boldsymbol{\beta}$ and γ, both entering in a non-linear manner and interacting with each other. This follows from the fact that $\vartheta = \exp(\boldsymbol{x}'\boldsymbol{\beta})$ and $\mu = \boldsymbol{x}'\boldsymbol{\gamma}$.

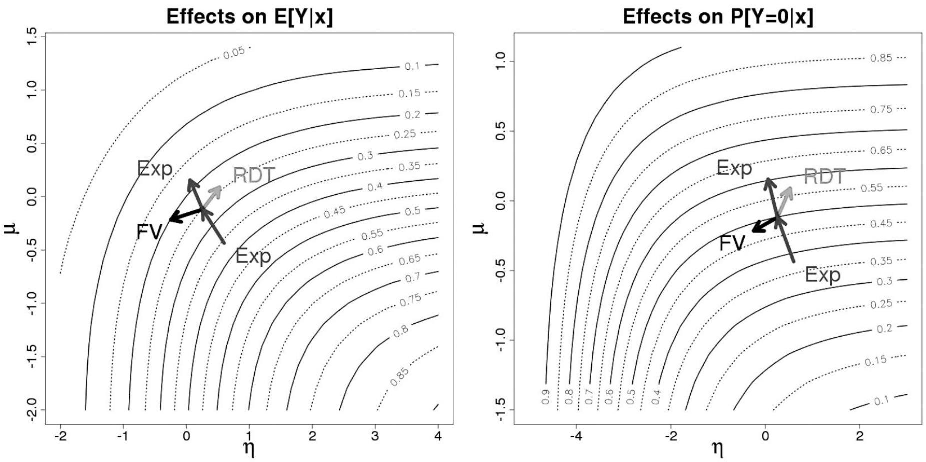

FIGURE 3: Illustration of effects of main covariates for the zero-inflated model. On the left hand side, a contour plot of the conditional expectation, $E[Y|\boldsymbol{x}]$, as a function of the two linear predictors $\eta = \boldsymbol{x}'\boldsymbol{\beta}$ and $\mu = \boldsymbol{x}'\boldsymbol{\gamma}$ is shown. On the right hand side, the same contour plot is shown for the probability of being zero, $P[Y = 0|\boldsymbol{x}]$. The arrows represent the effects of changing covariates. For the two continuous covariates face value (FV) and relative default time (RDT), the arrows are obtained by increasing the variables by one standard deviation from their mean. For the factorial variable experience (Exp), the two arrows indicate the changes when moving from the lowest level to the middle one and then to the highest level.

Because of this we came to the conclusion that interpretation is best done in a graphical way. This is done as described in the following.

In Figure 3, contour plots of the conditional expectation, $E[Y|\boldsymbol{x}]$, and the probability of being zero, $P[Y = 0|\boldsymbol{x}]$, for the fitted zero-inflated model are shown. Contour levels are obtained with respect to varying values of the two linear predictors $\eta = \boldsymbol{x}'\boldsymbol{\beta}$ and $\mu = \boldsymbol{x}'\boldsymbol{\gamma}$. The arrows represent the effects of the covariates. The middle point of the arrows are the levels of $E[Y|\boldsymbol{x}]$ and $P[Y = 0|\boldsymbol{x}]$, respectively, attained when taking all continuous covariates at their mean and the categorical variables at their most frequent level. We focus on the three variables face value (FV), relative default time (RDT), and experience (Exp) since these are believed to be the most important variables from a practical point of view. Interpretation for the other covariates is analogous. For the two continuous covariates face value (FV) and relative default time (RDT), the blue and red arrows in Figure 3 are obtained by increasing the variables by one standard deviation from their mean. For the categorical variable experience (Exp), the green arrows illustrate the changes in $E[Y|\boldsymbol{x}]$ and $P[Y = 0|\boldsymbol{x}]$ when moving from the lowest level to the middle one and then to the highest level of experience.

Concerning the conditional expectation, the blue arrow of the FV shows that an increase of FV by one standard deviation leads to an increase in $E[Y|\boldsymbol{x}]$ by about 0.05. RDT, on the other hand, has virtually no effect on the mean. Even though both linear predictors change considerably when increasing RDT, the change is along a contour level and has no effect on the value of $E[Y|\boldsymbol{x}]$.

Concerning the experience, we observe strong effects when going from low experience to middle and high, with a total decrease of about 0.17.

For the probability of being zero, the picture is slightly different. FV has only a small effect on $P[Y = 0|\boldsymbol{x}]$, whereas increasing RDT by one standard deviation results in an increase of about 6% in $P[Y = 0|\boldsymbol{x}]$. Experience again has a strong effect. $P[Y = 0|\boldsymbol{x}]$ increases by more than 20% when going from low to high experience.

5. Conclusion

Three special regression models for fractional response variables that attain their boundaries frequently were presented. The first model determines the distribution of the values between the limits and the frequency of the limiting values in a parsimonious way. Two extensions of this model to cover cases in which the frequencies of the limits do not follow this parsimonious description were introduced as well. The models were applied to a LGD dataset from insurance. They were found to fit the data in a specific insurance application better than other popular parametric models.

Acknowledgments

We thank Hans-Rudolf Künsch for helpful comments and discussions. We would also like to thank two anonymous referees for their insightful comments and suggestions.

APPENDIX

A. Proof of Lemma 2.1

Firstly, a censored gamma distribution with density as in (6) has expectation

$$\begin{aligned} E[Y|\boldsymbol{x}] &= 0 \cdot G_{\alpha,\vartheta}(\xi) + \int_0^1 y g_{\alpha,\vartheta}(y+\xi)\,dy + 1 \cdot (1 - G_{\alpha,\vartheta}(1+\xi)) \\ &= \int_{\xi}^{1+\xi} (z-\xi) g_{\alpha,\vartheta}(z)\,dz + (1 - G_{\alpha,\vartheta}(1+\xi)) \\ &= \alpha\vartheta(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)) + \xi G_{\alpha,\vartheta}(\xi) \\ &\quad - \xi G_{\alpha,\vartheta}(1+\xi) + (1 - G_{\alpha,\vartheta}(1+\xi)) \\ &= \alpha\vartheta(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi)) \\ &\quad + (1+\xi)\big(1 - G_{\alpha,\vartheta}(1+\xi)\big) - \xi\big(1 - G_{\alpha,\vartheta}(\xi)\big), \end{aligned} \tag{41}$$

where in the third line we have used the identity (47) given in the supplementary material.

Secondly, for a continuous x_j, using

$$\frac{\partial G_{\alpha,\vartheta}(\xi)}{\partial \vartheta} = \frac{\partial G_{\alpha,1}(\xi/\vartheta)}{\partial \vartheta} = -\frac{\xi}{\vartheta^2} g_{\alpha,1}\left(\frac{\xi}{\vartheta}\right) = -\frac{\xi}{\vartheta} g_{\alpha,\vartheta}(\xi), \tag{42}$$

or

$$\frac{\partial G_{\alpha,\vartheta}(\xi)}{\partial \vartheta} = -\alpha g_{\alpha+1,\vartheta}(\xi), \tag{43}$$

and the fact that

$$\frac{\partial \vartheta}{\partial x_j} = \vartheta \beta_j,$$

we can compute the partial derivatives of $E[Y|\boldsymbol{x}]$ with respect to x_j as

$$\begin{aligned}\frac{\partial E[Y|\boldsymbol{x}]}{\partial x_j} &= -\xi \alpha g_{\alpha+1,\vartheta}(\xi)\vartheta\beta_j + \alpha\vartheta\beta_j(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi))\\ &\quad - \alpha\vartheta\frac{1+\xi}{\vartheta} g_{\alpha+1,\vartheta}(1+\xi)\vartheta\beta_j + \alpha\vartheta\frac{\xi}{\vartheta} g_{\alpha+1,\vartheta}(\xi)\vartheta\beta_j\\ &\quad + (1+\xi)\alpha g_{\alpha+1,\vartheta}(1+\xi)\vartheta\beta_j\\ &= \alpha\vartheta(G_{\alpha+1,\vartheta}(1+\xi) - G_{\alpha+1,\vartheta}(\xi))\beta_j.\end{aligned} \tag{44}$$

REFERENCES

ABRAMOWITZ, M. and STEGUN, I.A. (1964) *Handbook of Mathematical Functions*. Dover Publications, New York.

AITCHISON, J. (1955) On the distribution of a positive random variable having a discrete probability mass at the origin. *J. Amer. Statist. Assoc*, **50**: 901-908.

AMEMIYA, T. (1985) *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts.

ARABMAZAR, A. and SCHMIDT, P. (1982) An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica*, **50(4)**: 1055-1063.

AZZALINI, A. and CAPITANIO, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65(2)**: 367-389.

BARDOSSY, A. and PLATE, A.J. (1992) Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, **28(5)**: 1247-1259.

BREEN, R. (1996) *Regression Models: Censored, Sample Selected, or Truncated Data*. Sage Publications, Thousand Oaks.

CHAMBERS, J.M. and HASTIE, T.J. (1992) *Statistical Models in S*. Wadsworth & Brooks/Cole.

CHEN, S. and KHAN, S. (2001) Semiparametric estimation of a partially linear censored regression model. *Econometric Theory*, **17(03)**: 567-590.

COOK, D.O., KIESCHNICK, R. and MCCULLOUGH, B.D. (2008) Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance*, **15(5)**: 860-867.

COUTURIER, D.L. and VICTORIA-FESER, M.-P. (2010) Zero-inflated truncated generalized pareto distribution for the analysis of radio audience data. *The Annals of Applied Statistics*, **4(4)**: 1824-1846.

CRAGG, J.G. (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, **39(5)**: 829-44.

CREPON, B. and DUGUET, E. (1997) Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *Journal of Econometrics*, **79(2)**: 355-378.

DEATON, A. and IRISH, M. (1984) Statistical models for zero expenditures in household budgets. *Journal of Public Economics*, **23(1-2)**: 59-80.

DEMPSTER, A.P., LAIRD, N.M. and RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39(1)**: 1-38.

FAHRMEIR, L. and TUTZ, G. (2001) *Multivariate statistical modelling based on generalized linear models*. Springer Series in Statistics. Springer-Verlag, New York.

FERRARI, S. and CRIBARI-NETO, F. (2004) Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31(7)**: 799-815.

GOLDBERGER, A.S. (1964) *Economic Theory*. Wiley, New York.

GOURIEROUX, C., MONFORT, A. and TROGNON, A. (1984) Pseudo maximum likelihood methods: Theory. *Econometrica*, **52(3)**: 681-700.

GURMU, S. (1997) Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *Journal of Applied Econometrics*, **12**: 225-242.

GURMU, S. and TRIVEDI, P.K. (1996) Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, **14(4)**: 469-77.

HECKMAN, J.J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, **5(4)**: 120-137.

HEILBRON, D.C. (1994) Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, **36**: 531-547.

KHAN, S. and POWELL, J.L. (2001) Two-step estimation of semiparametric censored regression models. *Journal of Econometrics*, **103(1-2)**: 73-110.

KOENKER, R. (2005) *Quantile Regression,* volume 1. Cambridge University Press, Cambridge University Press, 40 West 20th Street, New York.

LAMBERT, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**: 1-14.

LONG, J.S. (1997) *Regression Models for Categorical and Limited Dependent Variables*. Advances quantiative techniques in the social sciences; v. 7. SAGE Publications, Inc., Thousand Oaks, California 91320.

MADDALA, G.S. (1983) *Limited-dependent and qualitative variables in econometrics*, volume 3 of *Econometric Society Monographs in Quantitative Economics*. Cambridge University Press, Cambridge.

MADDALA, G.S. and NELSON, F.D. (1975) *Specification errors in limited dependent variable models*. NBER Working Papers 0096, National Bureau of Economic Research, Inc.

MCCULLAGH, P. and NELDER, J.A. (1983) *Generalized linear models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

MIAOU, S.-P. (1994) The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, **26**: 471-482.

MULLAHY, J. (1986) Specification and testing of some modified count data models. *Journal of Econometrics*, **33(3)**: 341-365.

PAOLINO, P. (2001) Maximum likelihood estimation of models with Beta-distributed dependent variables. *Political Analysis*, **9(4)**: 325-346.

PAPKE, L.E. and WOOLDRIDGE, J.M. (1996) Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, **11(6)**: 619-632.

PAPKE, L.E. and WOOLDRIDGE, J.M. (2008) Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, **145(1-2)**: 121-133.

PIESSENS, R., DEDONCKER-KAPENGA, E., UBERHUBER, C. and KAHANER, D. (1983) *Quadpack. A subroutine package for automatic integration*. Springer Series in Computational Mathematics, Volume 1. Springer-Verlag, New-York, 1983.

POWELL, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, **25(3)**: 303-325.

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J.M.R. (2011) Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, **25(1)**: 19-68.

Ramalho, J.J.S. and Vidigal da Silva, J. (2009) A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance*, **9(5)**: 621-636.

Ridout, M., Demetrio, C.G.B. and Hinde, J. (1998) Models for count data with many zeros. In *Proceedings of the XlXth International Biometrics Conference*, pages 179-190, Cape Town, December 1998.

Rosett, R.N. and Nelson, F.D. (1975) Estimation of the two-limit probit regression model. *Econometrica*, **43(1)**: 141-46.

Sanso, B. and Guenni, L. (2004) A Bayesian approach to compare observed rainfall data to deterministic simulations. *Environmetrics*, **15(6)**: 597-612.

Shonkwiler, J.S. and Shaw, W.D. (1996) Hurdle count-data models in recreation demand analysis. *Journal of Agricultural and Resource Economics*, **21(02)**: 210-219.

Tobin, H. (1958) Estimation of relationships for limited dependent variables. *Econometrica*, **26**: 24-36.

Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57(2)**: 307-333.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F. and Lindenmayer, D.B. (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, **88(1-3)**: 297-308.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, 1st edition. MIT Press, Cambridge.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd edition. MIT Press, Cambridge.

Fabio Sigrist (corresponding author)
Seminar for Statistics, Department of Mathematics
ETH Zurich
Rämistrasse 110
CH-8092 Zurich
Switzerland
E-Mail: sigrist@stat.math.ethz.ch

Werner A. Stahel
Seminar for Statistics, Department of Mathematics
ETH Zurich
Rämistrasse 110
CH-8092 Zurich
Switzerland
E-Mail: stahel@stat.math.ethz.ch

SUPPLEMENTARY MATERIAL

S.1. Fisher Information Matrix for the Censored Gamma Model

In the following derivations, we will often use some identities and results on integrals that we list in Section S.3 below.

With (16), it follows that

$$
\begin{aligned}
&E_\theta\left[\frac{\partial \ell_i}{\partial \alpha'}\frac{\partial \ell_i}{\partial \alpha'}\right]\\
&= E_\theta\left[\left(\frac{\alpha}{G_{\alpha,\vartheta_i}(\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi) + H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)\mathbf{1}_{\{y_i=0\}}\right)^2\right]\\
&+ E_\theta\left[\left(\alpha(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i+\xi))\,\mathbf{1}_{\{0<y_i<1\}}\right)^2\right]\\
&+ E_\theta\left[\left(-\frac{\alpha}{1-G_{\alpha,\vartheta_i}(1+\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi) + H_\alpha^{(1)}\left(0,\frac{1+\xi}{\vartheta_i}\right)\right)\mathbf{1}_{\{y_i=1\}}\right)^2\right]\\
&= \left(\frac{\alpha}{G_{\alpha,\vartheta_i}(\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi) + H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)\right)^2\cdot G_{\alpha,\vartheta_i}(\xi)\\
&+ \int_0^1\left(\alpha\left(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i+\xi)\right)\right)^2 g_{\alpha,\vartheta_i}(y_i+\xi)\,dy_i\\
&+ \left(\frac{\alpha}{1-G_{\alpha,\vartheta_i}(1+\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi) + H_\alpha^{(1)}\left(0,\frac{1+\xi}{\vartheta_i}\right)\right)\right)^2\cdot(1-G_{\alpha,\vartheta_i}(1+\xi)).
\end{aligned}
$$

Using (51) and (52), the middle summand of this expression is calculated as

$$
\begin{aligned}
&\int_0^1\left(\alpha\left(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i+\xi)\right)\right)^2 g_{\alpha,\vartheta_i}(y_i+\xi)\,dy_i\\
&= \alpha^2(\log(\vartheta_i)+\psi(\alpha))^2(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi))\\
&- 2\alpha^2(\log(\vartheta_i)+\psi(\alpha))\left(\log(\vartheta_i)(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) + H_\alpha^{(1)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\right)\\
&+ \alpha^2\log(\vartheta_i)^2(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) + 2\alpha^2\log(\vartheta_i)\,H_\alpha^{(1)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\\
&+ \alpha^2 H_\alpha^{(2)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\\
&= \alpha^2\psi(\alpha)^2(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) - 2\alpha^2\psi(\alpha)\,H_\alpha^{(1)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right) + \alpha^2 H_\alpha^{(2)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right).
\end{aligned}
$$

From this follows that

$$
\begin{aligned}
&E_\theta\left[\frac{\partial \ell_i}{\partial \alpha'}\frac{\partial \ell_i}{\partial \alpha'}\right]\\
&= \frac{\alpha^2}{G_{\alpha,\vartheta_i}(\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi) + H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)^2
\end{aligned}
$$

$$+ \alpha^2\Big(\psi(\alpha)^2(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) - 2\psi(\alpha)\,H_\alpha^{(1)}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big) + H_\alpha^{(2)}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big)\Big)$$
$$+ \frac{\alpha^2}{1 - G_{\alpha,\vartheta_i}(1+\xi)}\Big(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi) + H_\alpha^{(1)}\Big(0, \frac{1+\xi}{\vartheta_i}\Big)\Big)^2.$$

For the remaining entries of the Fisher Information Matrix, the calculation procedure is similar to the one made before. That is, the computation of each expectation can be split in to three terms of which the middle term, corresponding to the non-censored part of the model, requires more effort to compute. In the following, we therefore first calculate the corresponding middle term in each case.

With (47), (51), (53), and (45), we calculate

$$E_\theta\Big[\alpha\big(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i+\xi)\big)x_{ik}\Big(-\alpha + \frac{y_i+\xi}{\vartheta_i}\Big)\mathbf{1}_{\{0<y_i<1\}}\Big]$$
$$= \alpha^2 x_{ik}\log(\vartheta_i)(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) + \alpha^2 x_{ik}\,\psi(\alpha)(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi))$$
$$- \alpha^2 x_{ik}\log(\vartheta_i)(G_{\alpha+1,\vartheta_i}(1+\xi) - G_{\alpha+1,\vartheta_i}(\xi)) - \alpha^2 x_{ik}\,\psi(\alpha)(G_{\alpha+1,\vartheta_i}(1+\xi) - G_{\alpha+1,\vartheta_i}(\xi))$$
$$- \alpha^2 x_{ik}\log(\vartheta_i)(G_{\alpha,\vartheta_i}(1+\xi) - G_{\alpha,\vartheta_i}(\xi)) - \alpha^2 x_{ik}\,H_\alpha^{(1)}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big)$$
$$+ \alpha^2 x_{ik}\log(\vartheta_i)(G_{\alpha+1,\vartheta_i}(1+\xi) - G_{\alpha+1,\vartheta_i}(\xi)) + \alpha^2 x_{ik}\,H_{\alpha+1}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big)$$
$$= \alpha^2 x_{ik}\,\psi(\alpha)(G_{\alpha+1,\vartheta_i}(\xi) - G_{\alpha,\vartheta_i}(\xi) - G_{\alpha+1,\vartheta_i}(1+\xi) + G_{\alpha,\vartheta_i}(1+\xi))$$
$$+ \alpha^2 x_{ik}\Big(-H_\alpha^{(1)}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big) + H_{\alpha+1}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big)\Big)$$
$$= \alpha^2 x_{ik}\big(\psi(\alpha)\,\vartheta_i\,g_{\alpha+1,\vartheta_i}(1+\xi) - \psi(\alpha)\,\vartheta_i\,g_{\alpha+1,\vartheta_i}(\xi)\big)$$
$$- \alpha^2 x_{ik}\Big(H_\alpha^{(1)}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big) + H_{\alpha+1}\Big(\frac{\xi}{\vartheta_i}, \frac{1+\xi}{\vartheta_i}\Big)\Big).$$

Using this result, (16), and (17), we get

$$E_\theta\Big[\frac{\partial \ell_i}{\partial \alpha'}\frac{\partial \ell_i}{\partial \beta_k}\Big]$$
$$= E_\theta\Big[\frac{\alpha}{G_{\alpha,\vartheta_i}(\xi)}\Big(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi) + H_\alpha^{(1)}\Big(0, \frac{\xi}{\vartheta_i}\Big)\Big)\frac{-x_{ik}\,\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\,\mathbf{1}_{\{y_i=0\}}\Big]$$
$$+ E_\theta\Big[\alpha\big(-\log(\vartheta_i) - \psi(\alpha) + \log(y_i+\xi)\big)x_{ik}\Big(-\alpha + \frac{y_i+\xi}{\vartheta_i}\Big)\mathbf{1}_{\{0<y_i<1\}}\Big]$$
$$+ E_\theta\left[\frac{-\alpha\Big(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi) + H_\alpha^{(1)}\Big(0, \frac{1+\xi}{\vartheta_i}\Big)\Big)}{1 - G_{\alpha,\vartheta_i}(1+\xi)}\,\frac{x_{ik}(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)}{1 - G_{\alpha,\vartheta_i}(1+\xi)}\,\mathbf{1}_{\{y_i=1\}}\right]$$

$$
\begin{aligned}
&= -x_{ik}\frac{\alpha\xi\cdot g_{\alpha,\vartheta_i}(\xi)\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi)+H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)}{G_{\alpha,\vartheta_i}(\xi)}\\
&+x_{ik}\,\alpha^2\left(\psi(\alpha)\,\vartheta_i\,g_{\alpha+1,\vartheta_i}\left(\xi+1\right)-\psi(\alpha)\,\vartheta_i\,g_{\alpha+1,\vartheta_i}(\xi)\right)\\
&-x_{ik}\,\alpha^2\left(H_\alpha^{(1)}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)+H_{\alpha+1}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\right)\\
&-x_{ik}\frac{\alpha(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi)+H_\alpha^{(1)}\left(0,\frac{1+\xi}{\vartheta_i}\right)\right)}{1-G_{\alpha,\vartheta_i}(1+\xi)}.
\end{aligned}
$$

Next, with (47), (48), and (45), we calculate

$$
\begin{aligned}
&E_\theta\left[x_{ik}x_{il}\left(-\alpha+\frac{y_i+\xi}{\vartheta_i}\right)^2\mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad= x_{ik}x_{il}\alpha^2(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))-2\alpha^2x_{ik}x_{il}(G_{\alpha+1,\vartheta_i}(1+\xi)-G_{\alpha+1,\vartheta_i}(\xi))\\
&\quad+a(\alpha+1)\,x_{ik}x_{il}(G_{\alpha+2,\vartheta_i}(1+\xi)-G_{\alpha+2,\vartheta_i}(\xi))\\
&\quad= \alpha^2x_{ik}x_{il}\,\vartheta_i\left(g_{\alpha+1,\vartheta_i}(1+\xi)-g_{\alpha+1,\vartheta_i}(\xi)-g_{\alpha+2,\vartheta_i}(1+\xi)+g_{\alpha+2,\vartheta_i}(\xi)\right)\\
&\quad+\alpha x_{ik}x_{il}(G_{\alpha+2,\vartheta_i}(1+\xi)-G_{\alpha+2,\vartheta_i}(\xi)).
\end{aligned}
$$

Using this result and (17), we see that

$$
\begin{aligned}
&E_\theta\left[\frac{\partial\ell_i}{\partial\beta_k}\frac{\partial\ell_i}{\partial\beta_l}\right]\\
&= E_\theta\left[x_{ik}x_{il}\left(\frac{\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\right)^2\mathbf{1}_{\{y_i=0\}}\right]\\
&\quad+E_\theta\left[x_{ik}x_{il}\left(-\alpha+\frac{y_i+\xi}{\vartheta_i}\right)^2\mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad+E_\theta\left[x_{ik}x_{il}\left(\frac{(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\right)^2\mathbf{1}_{\{y_i=1\}}\right]\\
&= \alpha^2x_{ik}x_{il}\,\vartheta_i\left(g_{\alpha+1,\vartheta_i}(1+\xi)-g_{\alpha+1,\vartheta_i}(\xi)-g_{\alpha+2,\vartheta_i}(1+\xi)+g_{\alpha+2,\vartheta_i}(\xi)\right)\\
&\quad+x_{ik}x_{il}\left(\alpha(G_{\alpha+2,\vartheta_i}(1+\xi)-G_{\alpha+2,\vartheta_i}(\xi))+\frac{\left(\xi\cdot g_{\alpha,\vartheta_i}(\xi)\right)^2}{G_{\alpha,\vartheta_i}(\xi)}+\frac{\left((1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)\right)^2}{1-G_{\alpha,\vartheta_i}(1+\xi)}\right).
\end{aligned}
$$

Moreover, with (49), (51), (54), and (45), we get

$$
\begin{aligned}
&E_\theta\left[\alpha\big(-\log(\vartheta_i)-\psi(\alpha)+\log(y_i+\xi)\big)\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}}\right]\\
&=\frac{-\alpha\log(\vartheta_i)}{\vartheta_i}(G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\vartheta_i,\alpha-1}(\xi))-\frac{\alpha\psi(\alpha)}{\vartheta_i}(G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\vartheta_i,\alpha-1}(\xi))\\
&\quad+\frac{\alpha\log(\vartheta_i)}{\vartheta_i}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))+\frac{\alpha\psi(\alpha)}{\vartheta_i}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))\\
&\quad+\frac{\alpha\log(\vartheta_i)}{\vartheta_i}(G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\vartheta_i,\alpha-1}(\xi))+\frac{\alpha}{\vartheta_i}H_{\alpha-1}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\\
&\quad-\frac{\alpha\log(\vartheta_i)}{\vartheta_i}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))-\frac{\alpha}{\vartheta_i}H_{\alpha}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\\
&=\frac{\alpha\psi(\alpha)}{\vartheta_i}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\alpha,\vartheta_i}(\xi)+G_{\vartheta_i,\alpha-1}(\xi))\\
&\quad+\frac{\alpha}{\vartheta_i}\left(H_{\alpha-1}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)-H_{\alpha}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\right)\\
&=\alpha\psi(\alpha)\big(-g_{\alpha,\vartheta_i}(\xi+1)+g_{\alpha,\vartheta_i}(\xi)\big)+\frac{\alpha}{\vartheta_i}\left(H_{\alpha-1}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)-H_{\alpha}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\right).
\end{aligned}
$$

With this equation, (16), and (18), we calculate

$$
\begin{aligned}
&E_\theta\left[\frac{\partial\ell_i}{\partial\alpha'}\frac{\partial\ell_i}{\partial\xi}\right]\\
&=E_\theta\left[\frac{\alpha}{G_{\alpha,\vartheta_i}(\xi)}\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi)+H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)\frac{g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\mathbf{1}_{\{y_i=0\}}\right]\\
&\quad+E_\theta\left[\alpha\big(-\log(\vartheta_i)-\psi(\alpha)+\log(y_i+\xi)\big)\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad+E_\theta\left[\frac{-\alpha\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi)+H_\alpha^{(1)}\left(0,\frac{1+\xi}{\vartheta_i}\right)\right)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\frac{-g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\mathbf{1}_{\{y_i=1\}}\right]\\
&=\frac{\alpha\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(\xi)+H_\alpha^{(1)}\left(0,\frac{\xi}{\vartheta_i}\right)\right)g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\\
&\quad+\alpha\psi(\alpha)\big(-g_{\alpha,\vartheta_i}(\xi+1)+g_{\alpha,\vartheta_i}(\xi)\big)+\frac{\alpha}{\vartheta_i}\left(H_{\alpha-1}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)-H_{\alpha}\left(\frac{\xi}{\vartheta_i},\frac{1+\xi}{\vartheta_i}\right)\right)\\
&\quad+\frac{\alpha\left(-\psi(\alpha)\,G_{\alpha,\vartheta_i}(1+\xi)+H_\alpha^{(1)}\left(0,\frac{1+\xi}{\vartheta_i}\right)\right)g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}.
\end{aligned}
$$

With (47), (49), and (45), we calculate

$$
\begin{aligned}
&E_\theta\left[x_{ik}\left(-\alpha+\frac{y_i+\xi}{\vartheta_i}\right)\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}}\right]\\
&=\frac{-\alpha x_{ik}(G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\vartheta_i,\alpha-1}(\xi))}{\vartheta_i}+\frac{\alpha x_{ik}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))}{\vartheta_i}\\
&\quad+\frac{(\alpha-1)\,x_{ik}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))}{\vartheta_i}-\frac{\alpha x_{ik}(G_{\alpha+1,\vartheta_i}(1+\xi)-G_{\alpha+1,\vartheta_i}(\xi))}{\vartheta_i}\\
&=x_{ik}\alpha\big(-g_{\alpha,\vartheta_i}(1+\xi)+g_{\alpha,\vartheta_i}(\xi)+g_{\alpha+1,\vartheta_i}(1+\xi)-g_{\alpha+1,\vartheta_i}(\xi)\big)\\
&\quad-\frac{x_{ik}(G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi))}{\vartheta_i}.
\end{aligned}
$$

Using the above result, we have

$$
\begin{aligned}
&E_\theta\left[\frac{\partial\ell_i}{\partial\beta_k}\frac{\partial\ell_i}{\partial\xi}\right]\\
&=E_\theta\left[-x_{ik}\frac{\xi g_{\alpha,1}\left(\frac{\xi}{\vartheta_i}\right)}{G_{\alpha,\vartheta_i}(\xi)}\frac{g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\mathbf{1}_{\{y_i=0\}}\right]\\
&\quad+E_\theta\left[x_{ik}\left(-\alpha+\frac{y_i+\xi}{\vartheta_i}\right)\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)\mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad+E_\theta\left[x_{ik}\frac{(1+\xi)g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\frac{-g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\mathbf{1}_{\{y_i=1\}}\right]\\
&=x_{ik}\alpha(-g_{\alpha,\vartheta_i}(1+\xi)+g_{\alpha,\vartheta_i}(\xi)+g_{\alpha+1,\vartheta_i}(1+\xi)-g_{\alpha+1,\vartheta_i}(\xi))\\
&\quad+x_{ik}\left(-\frac{G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi)}{\vartheta_i}-\frac{\xi g_{\alpha,\vartheta_i}(\xi)^2}{G_{\alpha,\vartheta_i}(\xi)}-\frac{(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)^2}{1-G_{\alpha,\vartheta_i}(1+\xi)}\right).
\end{aligned}
$$

Next, with (49), (50), (45), we calculate

$$
\begin{aligned}
&E_\theta\left[\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)^2\mathbf{1}_{\{0<y_i<1\}}\right]\\
&=\frac{(\alpha-1)(G_{\vartheta_i,\alpha-2}(1+\xi)-G_{\vartheta_i,\alpha-2}(\xi))}{(\alpha-2)\,\vartheta_i^2}-2\,\frac{G_{\vartheta_i,\alpha-1}(1+\xi)-G_{\vartheta_i,\alpha-1}(\xi)}{\vartheta_i^2}\\
&+\frac{G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi)}{\vartheta_i^2}\\
&=\left(\frac{\vartheta_i(\alpha-1)^2-(\xi+1)(\alpha-3)}{\vartheta_i(\alpha-2)(\xi+1)}\right)g_{\alpha,\vartheta_i}(\xi+1)-\left(\frac{\vartheta_i(\alpha-1)^2-\xi(\alpha-3)}{\vartheta_i(\alpha-2)\xi}\right)g_{\alpha,\vartheta_i}(\xi)\\
&+\frac{G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi)}{(\alpha-2)\,\vartheta_i^2}.
\end{aligned}
$$

Finally, using this result, we have

$$
\begin{aligned}
&E_\theta\left[\frac{\partial \ell_i}{\partial \xi}\frac{\partial \ell_i}{\partial \xi}\right]\\
&= E_\theta\left[\left(\frac{g_{\alpha,\vartheta_i}(\xi)}{G_{\alpha,\vartheta_i}(\xi)}\right)^2 \mathbf{1}_{\{y_i=0\}}\right] + E_\theta\left[\left(\frac{\alpha-1}{y_i+\xi}-\frac{1}{\vartheta_i}\right)^2 \mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad + E_\theta\left[\left(\frac{-g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}\right)^2 \mathbf{1}_{\{y_i=1\}}\right]\\
&= \frac{g_{\alpha,\vartheta_i}(\xi)^2}{G_{\alpha,\vartheta_i}(\xi)} + \left(\frac{\vartheta_i(\alpha-1)^2-(\xi+1)(\alpha-3)}{\vartheta_i(\alpha-2)(\xi+1)}\right) g_{\alpha,\vartheta_i}(\xi+1)\\
&\quad - \left(\frac{\vartheta_i(\alpha-1)^2-\xi(\alpha-3)}{\vartheta_i(\alpha-2)\xi}\right) g_{\alpha,\vartheta_i}(\xi) + \frac{G_{\alpha,\vartheta_i}(1+\xi)-G_{\alpha,\vartheta_i}(\xi)}{(\alpha-2)\vartheta_i^2} + \frac{g_{\alpha,\vartheta_i}(1+\xi)^2}{1-G_{\alpha,\vartheta_i}(1+\xi)}.
\end{aligned}
$$

S.2. Fisher Information Matrix for the Two-tiered Gamma Model

First, with (47) and (48), we get

$$
\begin{aligned}
E_\theta\left[\frac{\partial \ell_i}{\partial \beta_k}\frac{\partial \ell_i}{\partial \beta_l}\right] &= E_\theta\left[x_{ik}x_{il}\left(\frac{y_i+\xi}{\vartheta_i}-\alpha-\frac{\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{\vartheta_i\cdot(1-G_{\alpha,\vartheta_i}(\xi))}\right)^2 \mathbf{1}_{\{0<y_i<1\}}\right]\\
&\quad + E_\theta\left[\frac{x_{ik}x_{il}}{\vartheta_i^2}\cdot\left(\frac{(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)}-\frac{\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{1-G_{\alpha,\vartheta_i}(\xi)}\right)^2 \mathbf{1}_{\{y_i=1\}}\right]\\
&= x_{ik}x_{il}\,\alpha\frac{(G_{\alpha+2,\vartheta_i}(1+\xi)-G_{\alpha+2,\vartheta_i}(\xi))(1-G_{\tilde\vartheta_i,\alpha}(\xi))}{1-G_{\alpha,\vartheta_i}(\xi)}\\
&\quad + x_{ik}x_{il}\,\frac{\alpha^2\left(\xi g_{\alpha+1,\vartheta_i}(\xi)-(1+\xi)g_{\alpha+1,1}\left(\frac{1+\xi}{\vartheta_i}\right)\right)(1-G_{\tilde\vartheta_i,\alpha}(\xi))}{(\alpha+1)(1-G_{\alpha,\vartheta_i}(\xi))}\\
&\quad + x_{ik}x_{il}\,\frac{\alpha\left((1+\xi)g_{\alpha,\vartheta_i}(1+\xi)-\xi g_{\alpha,\vartheta_i}(\xi)\right)(1-G_{\tilde\vartheta_i,\alpha}(\xi))}{1-G_{\alpha,\vartheta_i}(\xi)}\\
&\quad - x_{ik}x_{il}\,\frac{\xi^2 g_{\alpha,\vartheta_i}(\xi)^2(1-G_{\tilde\vartheta_i,\alpha}(\xi))}{(1-G_{\alpha,\vartheta_i}(\xi))^2}\\
&\quad + x_{ik}x_{il}\,\frac{(1+\xi)^2 g_{\alpha,\vartheta_i}(1+\xi)^2(1-G_{\tilde\vartheta_i,\alpha}(\xi))}{(1-G_{\alpha,\vartheta_i}(\xi))(1-G_{\alpha,\vartheta_i}(1+\xi))}
\end{aligned}
$$

Next, with (47) and the identity in (46), we get

$$
\begin{aligned}
E_\theta\left[\frac{\partial \ell_i}{\partial \beta_k}\frac{\partial \ell_i}{\partial \gamma_l}\right] =\ & E_\theta\left[x_{ik}x_{il}\left(\frac{y_i+\xi}{\vartheta_i} - \alpha - \frac{\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{1-G_{\alpha,\vartheta_i}(\xi)}\right)\frac{\xi\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{1-G_{\tilde{\vartheta}_i,\alpha}(\xi)}\mathbf{1}_{\{0<y_i<1\}}\right] \\
& + E_\theta\left[x_{ik}x_{il}\cdot\left(\frac{(1+\xi)\cdot g_{\alpha,\vartheta_i}(1+\xi)}{1-G_{\alpha,\vartheta_i}(1+\xi)} - \frac{\xi\cdot g_{\alpha,\vartheta_i}(\xi)}{1-G_{\alpha,\vartheta_i}(\xi)}\right)\frac{\xi\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{1-G_{\tilde{\vartheta}_i,\alpha}(\xi)}\mathbf{1}_{\{y_i=1\}}\right] \\
=\ & 0.
\end{aligned}
$$

Finally, we calculate

$$
\begin{aligned}
E_\theta\left[\frac{\partial \ell_i}{\partial \gamma_k}\frac{\partial \ell_i}{\partial \gamma_l}\right] =\ & E_\theta\left[x_{ik}x_{il}\left(\frac{\xi\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{\tilde{\vartheta}_i\cdot G_{\tilde{\vartheta}_i,\alpha}(\xi)}\right)^2\mathbf{1}_{\{y_i=0\}}\right] \\
& + E_\theta\left[x_{ik}x_{il}\left(\frac{\xi\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)}{\tilde{\vartheta}_i\cdot(1-G_{\tilde{\vartheta}_i,\alpha}(\xi))}\right)^2\left(\mathbf{1}_{\{0<y_i<1\}}+\mathbf{1}_{\{y_i=1\}}\right)\right] \\
=\ & x_{ik}x_{il}\frac{\xi^2\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)^2}{G_{\tilde{\vartheta}_i,\alpha}(\xi)} + x_{ik}x_{il}\frac{\xi^2\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)^2}{1-G_{\tilde{\vartheta}_i,\alpha}(\xi)} \\
=\ & x_{ik}x_{il}\frac{\xi^2\cdot g_{\alpha,\tilde{\vartheta}_i}(\xi)^2}{G_{\tilde{\vartheta}_i,\alpha}(\xi)\,(1-G_{\tilde{\vartheta}_i,\alpha}(\xi))}.
\end{aligned}
$$

S.3. Useful Identities and Integrals

By partial integration, we calculate

$$
\begin{aligned}
G_{\alpha+1,\vartheta}(\xi) =\ & \frac{1}{\vartheta^{\alpha+1}\Gamma(\alpha+1)}\int_0^h y^\alpha \exp(-y/\vartheta)\,dy \\
=\ & \frac{1}{\vartheta^{\alpha+1}\Gamma(\alpha+1)}\left(-\xi^\alpha s\exp(-\xi/\vartheta)\right) \\
& + \frac{1}{\vartheta^{\alpha+1}\Gamma(\alpha+1)}\int_0^h a y^{\alpha-1} s\exp(-y/\vartheta)\,dy \\
=\ & -\frac{1}{\Gamma(\alpha+1)}\left(\frac{\xi}{\vartheta}\right)^\alpha \exp(-\xi/\vartheta) + \frac{1}{\vartheta^\alpha\Gamma(\alpha)}\int_0^h y^{\alpha-1}\exp(-y/\vartheta)\,dy \\
=\ & -\vartheta g_{\alpha+1,\vartheta}(\xi) + G_{\alpha,\vartheta}(\xi).
\end{aligned}
$$

And from this follows

$$
G_{\alpha+1,\vartheta}(\xi) - G_{\alpha,\vartheta}(\xi) = -\vartheta\, g_{\alpha+1,\vartheta}(\xi) \tag{45}
$$

or

$$G_{\alpha+1,\vartheta}(\xi) - G_{\alpha,\vartheta}(\xi) = -\frac{\xi}{\alpha} g_{\alpha,\vartheta}(\xi). \tag{46}$$

For $0 \leq l < u$, the following equations hold true.

$$\int_l^u y g_{\alpha,\vartheta}(y)\, dy = \alpha\vartheta(G_{\alpha+1,\vartheta}(u) - G_{\alpha+1,\vartheta}(l)). \tag{47}$$

$$\int_l^u y^2 g_{\alpha,\vartheta}(y)\, dy = \vartheta^2 a(\alpha+1)(G_{\alpha+2,\vartheta}(u) - G_{\alpha+2,\vartheta}(l)). \tag{48}$$

$$\int_l^u \frac{1}{y} g_{\alpha,\vartheta}(y)\, dy = \frac{1}{(\alpha-1)\vartheta}(G_{\alpha-1,\vartheta}(u) - G_{\alpha-1,\vartheta}(l)). \tag{49}$$

$$\int_l^u \frac{1}{y^2} g_{\alpha,\vartheta}(y)\, dy = \frac{1}{(\alpha-1)(\alpha-2)\vartheta^2}(G_{\alpha-2,\vartheta}(u) - G_{\alpha-2,\vartheta}(l)). \tag{50}$$

$$\int_l^u \log(y) g_{\alpha,\vartheta}(y)\, dy = \log(\vartheta)(G_{\alpha,\vartheta}(u) - G_{\alpha,\vartheta}(l)) + H_\alpha^{(1)}\left(\frac{l}{\vartheta}, \frac{u}{\vartheta}\right). \tag{51}$$

$$\begin{aligned}\int_l^u \log(y)^2 g_{\alpha,\vartheta}(y)\, dy &= \log(\vartheta)^2 (G_{\alpha,\vartheta}(u) - G_{\alpha,\vartheta}(l)) \\ &\quad + 2\log(\vartheta) H_\alpha^{(1)}\left(\frac{l}{\vartheta}, \frac{u}{\vartheta}\right) + H_\alpha^{(2)}\left(\frac{l}{\vartheta}, \frac{u}{\vartheta}\right).\end{aligned} \tag{52}$$

$$\begin{aligned}\int_l^u y\log(y) g_{\alpha,\vartheta}(y)\, dy &= \alpha\vartheta\log(\vartheta)(G_{\alpha+1,\vartheta}(u) - G_{\alpha+1,\vartheta}(l)) \\ &\quad + \alpha\vartheta H_{\alpha+1}^{(1)}\left(\frac{l}{\vartheta}, \frac{u}{\vartheta}\right).\end{aligned} \tag{53}$$

$$\begin{aligned}\int_l^u \frac{\log(y)}{y} g_{\alpha,\vartheta}(y)\, dy &= \frac{1}{(\alpha-1)\vartheta}\log(\vartheta)(G_{\alpha-1,\vartheta}(u) - G_{\alpha-1,\vartheta}(l)) \\ &\quad + \frac{1}{(\alpha-1)\vartheta} H_{\alpha-1}^{(1)}\left(\frac{l}{\vartheta}, \frac{u}{\vartheta}\right).\end{aligned} \tag{54}$$

S.4. Descriptive statistics for covariates

TABLE 2

DESCRIPTIVE STATISTICS FOR COVARIATES.
FOR CATEGORICAL VARIABLES, THE FREQUENCY (IN %) OF THE LEVELS ARE GIVEN.

	Mean	Standard Deviation		
RDT	0.65	0.30		
Face Value (log)	3.84	0.54		
Insured Fraction	0.06	0.03		
	Low / Small	Mid / Medium	High / Large	
Experience	15.52	55.38	29.10	
Size	85.06	9.95	4.98	
	Maintenance	Hybrid	Performance	Other
Type	88.05	6.85	4.42	0.67

S.5. Additional Plots Illustrating Other Fitted Models

Additionaly, two different types of models have been fitted two the data. First, the two-limit version of the normal Tobit model and its corresponding two-tiered and zero-inflated extensions. Further, we fitted models using the skewed t-distribtion (Azzalini and Capitanio (2003)) where, in each model, the shifted Gamma distribution is replaced by a skewed t-distribution. The degrees of freedom were chosen to be 1 since this provided the best fit in general.

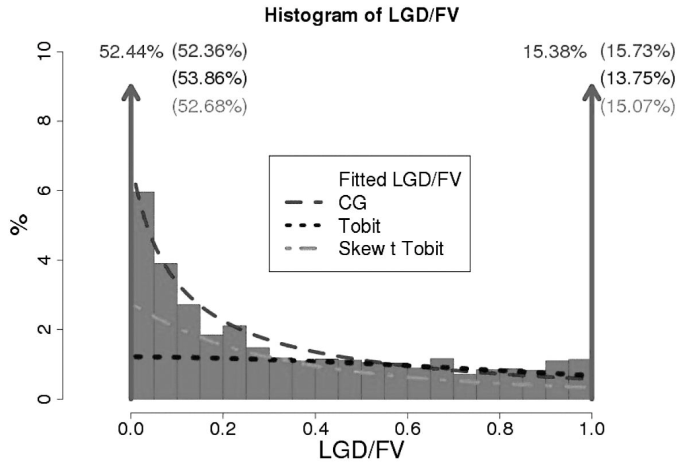

FIGURE 4: Comparison of fitted censored gamma, normal Tobit, and skew t Tobit models with no covariates. The numbers above the blue arrows represent the percentage of LGD/FV's being exactly zero or one, respectively. In parentheses are the corresponding numbers as predicted by the models.

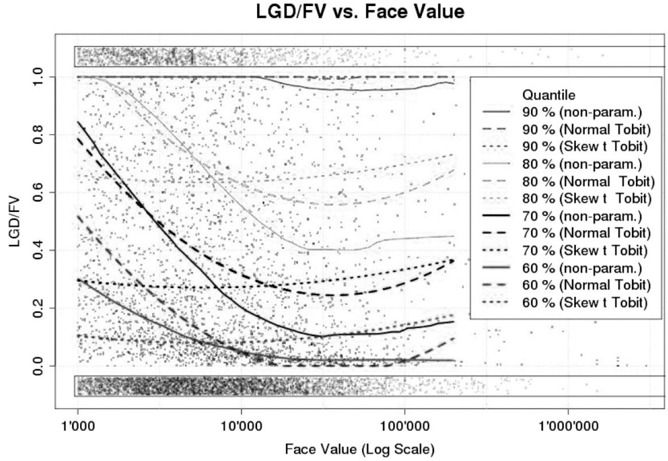

FIGURE 5: Scatter plot of LGD/FV versus face value (on a logarithmic scale). The jittered points in the bars below 0.0 and above 1.0 represent bonds with LGD/FV being exactly zero and one, respectively. The colored solid lines are non-parametrically fitted quantiles and the mean. The dashed and dotted lines represent quantiles of the fitted normal Tobit model and the skew t Tobit model, respetively. Logarithmic and squared logarithmic face value are taken as covariates.

TABLE 3

Fitted censored, two-tiered, and zero-inflated normal Tobit models including all covariates. Codes for significance levels: '***': $p < 0.001$, '**': $0.001 \leq p < 0.01$, '*': $0.01 \leq p < 0.05$, '.': $0.05 \leq p < 0.1$.

Covariate	Model	Censored		Two-Tiered				Zero-Inflated			
		Coef	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.
Intercept		0.86	0.13***	15.	0.030***	15.	0.030***	2.1	0.19***	2.1	0.19***
RDT	Lin	−0.19	0.044***	1.5	0.049***	0.41	0.33	−0.042	0.071	4.5	1.1***
	Quad	−0.27	0.15 ·	8.3	0.087***	−0.70	0.11***	0.33	0.24	0.59	0.20**
Experience	Lin	−0.40	0.028***	−2.0	0.0041***	−1.6	0.38***	−0.32	0.043***	2.3	0.65***
	Quad	0.036	0.020 ·	0.29	0.0091***	−0.87	0.073***	0.013	0.027	0.36	0.17*
Size	Lin	0.40	0.16*	2.8	0.051***	0.036	0.053	0.23	0.21	−0.094	0.092
	Quad	0.51	0.10***	0.10	0.0098***	0.91	0.40*	0.049	0.13	−0.27	0.51
Face Value	Lin	−0.22	0.026***	−5.4	0.0017***	1.4	0.25***	−0.50	0.042***	−1.5	0.31***
	Quad	0.22	0.025***	0.27	0.0021***	−0.047	0.068	0.26	0.034***	−1.5	0.31***
Type	Hybrid	1.4	0.47**	8.9	0.021***	0.64	0.075***	1.0	0.46*	−0.78	0.32*
	Performance	−0.036	0.052	0.96	1.7	3.3	1.5*	0.0095	0.061	−3.5	2.7
	Other	−0.019	0.071	2.4	0.031***	−0.16	0.14	0.035	0.10	0.19	0.31
Ins. Frac.		0.29	0.19	8.7	0.99***	−0.28	0.19	0.76	0.22***	0.26	0.37
		Value	Std. Err.		Value	Std. Err.			Value	Std. Err.	
log(σ)		−0.040	0.016		0.81	0.0014			−0.15	0.022	
Log-Likelihood		−8241.6		−7864.4				−8169			
AIC		16511.2		15780.8				16390			

TABLE 4

FITTED CENSORED, TWO-TIERED, AND ZERO-INFLATED SKEW t (df = 1) TOBIT MODELS INCLUDING ALL COVARIATES. CODES FOR SIGNIFICANCE LEVELS: '***': $p < 0.001$, '**': $0.001 \leq p < 0.01$, '*': $0.01 \leq p < 0.05$, '.': $0.05 \leq p < 0.1$.

Model		Censored		Two-Tiered				Zero-Inflated			
Covariate		Coef	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.	Coef(β)	Std. Err.	Coef(γ)	Std. Err.
Intercept		−0.38	0.056	−1.9	0.10	0.015	0.0090 ·	−0.090	0.075	−0.49	0.33
RDT	Lin	−0.15	0.023***	−0.15	0.037***	−0.010	0.0025***	−0.19	0.041***	−0.17	0.16
	Quad	−0.38	0.073***	−0.36	0.12**	−0.026	0.0080**	−0.39	0.14**	0.23	0.54
Experience	Lin	−0.12	0.013***	0.20	0.034***	−0.016	0.0037***	0.050	0.026 ·	1.3	0.21***
	Quad	−0.013	0.0094	−0.060	0.022**	0.0020	0.0011 ·	−0.0017	0.023	−0.28	0.15 ·
Size	Lin	0.26	0.10**	−0.20	0.13	0.012	0.0060 ·	0.42	0.15**	0.48	0.41
	Quad	0.33	0.069***	0.0064	0.073	0.021	0.0048***	0.33	0.10**	−0.11	0.30
Face Value	Lin	0.038	0.010***	0.40	0.019***	0.0024	0.00079**	0.030	0.0074***	0.040	0.066
	Quad	0.042	0.0064***	−0.12	0.011***	0.011	0.0028***	0.0069	0.016	−0.44	0.080***
Type	Hybrid	0.56	0.11***	−2.1	0.48***	0.066	0.033*	0.020	0.22	−1.8	1.1 ·
	Performance	−0.038	0.021 ·	−0.0033	0.026	−0.0038	0.0024	−0.034	0.037	−0.038	0.19
	Other	−0.12	0.0099***	−0.050	0.031	−0.0070	0.0033*	−0.00075	0.062	0.25	0.21
Ins. Frac.		−0.13	0.085	−1.1	0.64 ·	−0.013	0.0098	0.50	0.084***	1.1	0.34**
		Value	Std. Err.		Value	Std. Err.			Value	Std. Err.	
Skew t Par.	ν	1			1				1		
	$\log(\sigma)$	−1.1	0.028		−3.5	0.22			−0.90	0.033	
	α	30.	23.		−1.0	0.31			38.	27.	
Log-Likelihood		−8019		−7692.4				−7964.4			
AIC		16067.9		15440.7				15984.8			

https://doi.org/10.2143/AST.41.2.2136992